High-spatial-resolution transcriptome profiling reveals uncharacterized regulatory complexity underlying cambial growth and wood formation in *Populus tremula*

David Sundell[1*], Nathaniel R. Street[1*], Manoj Kumar[2*], Ewa J. Mellerowicz[2], Melis Kucukoglu[2], Christoffer Johnsson[2], Vikash Kumar[2], Chanaka Mannapperuma[1], Ove Nilsson[2], Hannele Tuominen[1], Edouard Pesquet[1,3], Urs Fischer[2], Totte Niittyla[2], Björn Sundberg[2] and Torgeir R. Hvidsten[1,4,†]

[1]Umeå Plant Science Center, Department of Plant Physiology, Umeå University, Umeå, Sweden.
[2]Umeå Plant Science Centre, Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, Umeå, Sweden.
[3]Department of Ecology, Environment and Plant Sciences, Stockholm University, Sweden.
[4]Department of Chemistry, Biotechnology and Food Sciences, Norwegian University of Life Sciences, Ås, Norway.

[*]Equal contributors
[†]Correspondence: torgeir.hvidsten@umu.se

Abstract

Trees represent the largest terrestrial carbon sink and a renewable source of ligno-cellulose. There is significant scope for yield and quality improvement in these largely undomesticated species, however, efforts to engineer new, elite varieties are constrained by the lack of a comprehensive understanding of the transcriptional network underlying cambial growth and wood formation. We generated RNA Sequencing transcriptome data for four mature, wild-growing aspens (*Populus tremula*) from high-spatial-resolution tangential cryosection series spanning the secondary phloem, vascular cambium, expanding and secondary cell wall forming xylem cells, cell death zone and the previous year's annual ring. The transcriptome comprised 28,294 expressed, previously annotated protein-coding genes, 78 novel protein-coding genes and 567 long intergenic non-coding RNAs. Most paralogs originating from the *Salicaceae* whole genome duplication had diverged expression, with the notable exception of those with high expression during secondary cell wall deposition. We performed co-expression network analysis to identify central transcriptional modules and associated several of these with known biological processes. This revealed previously uncharacterized complexity underlying the regulation of cambial growth and wood formation, with modules forming a continuum of activated processes across the tissues. The high spatial resolution suggested novel roles for known genes involved in xylan and cellulose biosynthesis, regulators of xylem vessel and fiber differentiation and components of lignification. The associated web resource (AspWood, http://aspwood.popgenie.org) integrates the data within a set of interactive tools for exploring the co-expression network of cambial growth and wood formation.

Keywords
Cambium growth, wood formation, transcriptomics, co-expression networks, paralogues, cell wall

Introduction

Trees dominate forest ecosystems, with the majority of biomass residing in the wood of stems, branches and roots. Wood formation is initiated in the vascular cambium meristem (hereafter, cambium), which has been described in many studies (Barnett, 1981; Larson, 1994; Mellerowicz et al., 2001; Johnsson and Fischer, 2016). The cambium forms a cylindrical shell of dividing cells within the stem. Inwards it forms secondary xylem (wood) and outwards secondary phloem cells are added to the growing stem. The cambium consists of stem cells (also referred to as initials) and their dividing derivatives (also referred to as xylem and phloem mother cells). The stem cells retain the capacity to divide over long periods of time (stem cell maintenance), whereas derivative cells divide over a few cell cycles. Before terminal differentiation into specialized cell types, all xylem and phloem cells undergo initial cell expansion and primary cell wall (PCW) biosynthesis (Johnsson and Fischer, 2016). In most broad-leaved trees (hardwoods), cambial derivatives differentiate into four major wood cell types; fibers that provide structural support, vessel elements for water and mineral transport, axial parenchyma cells for storage, and ray cells involved in radial transport and storage of photosynthates. Formation of the secondary cell wall (SCW) and lignification occur in all xylem cell types. Lignin composition and quantity, however, differs between the primary and secondary cell walls as well as between vessel elements, fibers and ray cells (Barros et al., 2015). Vessel elements and fibers undergo both programmed cell death (PCD), with death occurring earlier in vessel elements than in fibers (Courtois-Moreau et al., 2009), while ray cells remain alive for several years (Nakaba et al., 2012). The duration of SCW thickening and lignification is difficult to define, and has even been shown to progress *post-mortem* for lignin in vessel (Pesquet et al., 2013). Distinct visible markers that would identify sub-stages of the differentiation processes are therefore not available. Derivatives at the outer cambial face differentiate into the different cell types forming the phloem including sieve tube cells involved in long-range transport of photosynthates, ray and axial parenchyma cells, including companion cells, and phloem fibers with thickened and lignified cell walls (Evert and Eichhorn, 2006).

Transcriptomics and genetic analyses have been used to enhance our understanding of the underlying molecular mechanisms of secondary growth using *A. thaliana* as a model system (Ruzicka et al., 2015; Fukuda, 2016). However, the small size and limited extent of secondary growth of *A. thaliana* severely limits the spatial resolution at which SCW formation and vascular development can be assayed. Additionally, the xylem fibers of *A. thaliana* do not fully mature but stay alive even during prolonged growth, at least in greenhouse conditions (Bollhoner et al., 2012). Use of woody tree systems, such as *Populus spp.,* overcome these limitations and transcriptomic analyses in such species have enhanced our understanding of the molecular mechanisms underlying cambial growth and wood formation. The first study of the wood formation transcriptome was performed by Hertzberg et al. (2001) using cDNA microarrays covering less than 10% of the *Populus* gene space (2,995 probes). This study utilized tangential cryosectioning, as first described in Uggla et al. (1996), to obtain transcriptomes

from different stages of wood development. Subsequently, a number of similar studies were performed using microarrays with higher gene coverage (Schrader et al., 2004; Aspeborg et al., 2005; Moreau et al., 2005; Courtois-Moreau et al., 2009) and, recently, RNA-Sequencing (RNA-Seq) (Immanen et al., 2016). However, earlier transcriptome studies of cambial growth and wood formation in *Populus* were limited to specific wood forming tissues using supervised approaches, i.e. samples were collected on the basis of visual anatomical assessment during sectioning.

The availability of RNA-Seq methodologies enables transcriptome profiling with higher dynamic range than was previously possible. Moreover, RNA-Seq does not require prior knowledge of gene models and can be used to identify effectively all transcribed loci in a sample (e.g. Goodwin et al. (2016)). The base-pair resolution of RNA-Seq additionally enables differentiation of expression from transcripts arising from highly sequence-similar paralogous genes or gene-family members. This is particularly pertinent for studies within the *Salicaceae* family, which underwent a relatively recent whole genome duplication (WGD) event, with about half of all genes still being part of a paralogous gene pair in the *P. trichocarpa* reference genome (Sterck et al., 2005; Tuskan et al., 2006; Goodstein et al., 2012). Concomitantly, tools for analyzing expression data, such as co-expression networks, represent increasingly powerful approaches for identifying central genes with important functional roles (e.g. Mutwil et al. (2011) and Netotea et al. (2014)).

Here, we employed RNA-Seq to assay gene expression across wood-forming tissues in aspen (*Populus tremula*). Cryosectioning was used to obtain a continuous sequence of samples extending from differentiated phloem to mature xylem with high spatial resolution (25-28 samples per replicate). This enabled a continuous and unsupervised analysis of transcriptional modules, in which they were assigned developmental context using marker genes with known activity in particular domains. We generated expression profiles for 28,294 previously annotated protein-coding genes in addition to 78 novel protein-coding genes and 567 long intergenic non-coding RNAs that we identified *de novo*. Co-expression network analysis was used to identify 41 transcriptional modules representing major events in cambial growth and wood formation. Several modules could be related to discrete and well described processes such as cell division, cell expansion and ligno-cellulose formation, but the functional role of many of the modules remained elusive. The high spatial resolution of the data also revealed potential novel roles of well-studied genes involved in cellulose and xylan biosynthesis, regulators of xylem vessel and fiber differentiation and lignification. Analyzing the regulatory fate of paralogs arising from the WGD showed that a majority displayed distinctly different expression profiles across wood formation. However, we found evidence indicating that paralogs with conserved expression may have been retained to achieve high expression during SCW formation. The data is publically available in the AspWood (Aspen Wood, http://aspwood.popgenie.org) interactive web resource, which enables genomics analysis of cambial growth and wood formation.

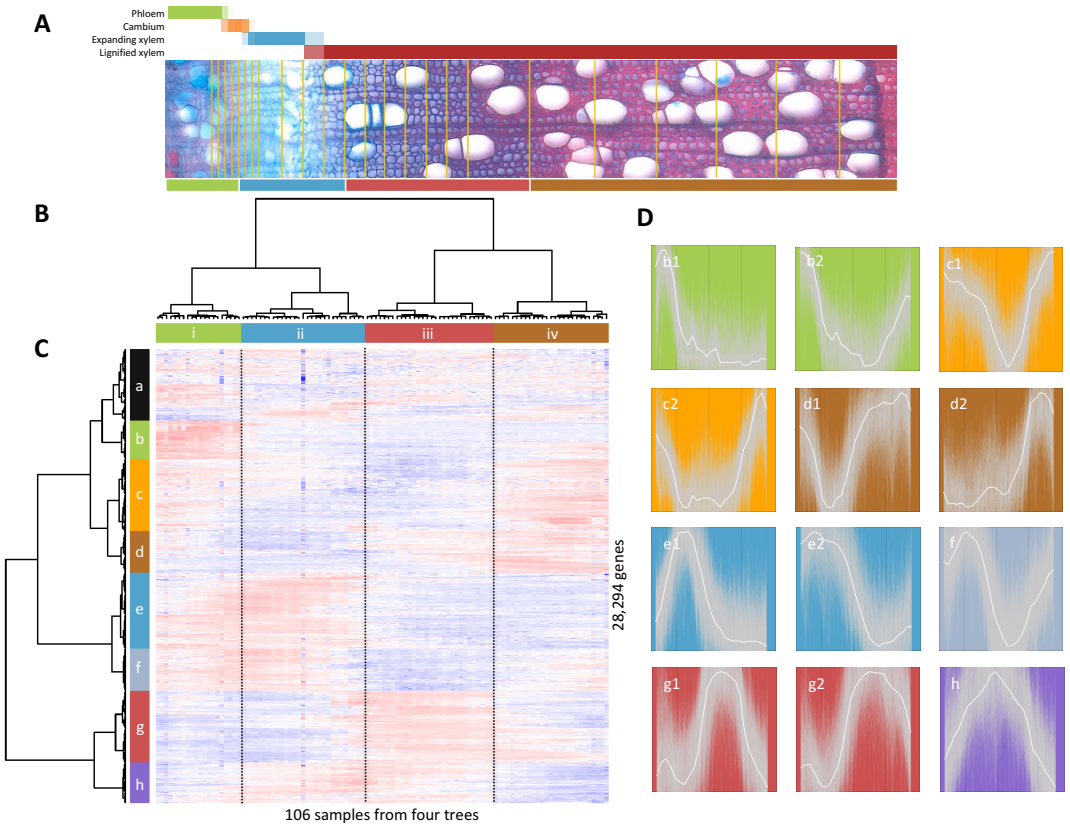

Figure 1. *Hierarchical clustering of samples and genes across developing xylem and phloem tissues. (A) Transverse cross-section image from one of the sampled trees (tree T1). The strategy for pooling the samples for RNA-Seq is visualized by overlaying the samples on the section (positions on the section are approximated). The color bar below the image shows four sample clusters identified by hierarchical clustering (see B). The color bar above the image shows the estimated tissue composition for each sample. (B) Hierarchical clustering of all 106 samples from the four replicate trees using mRNA expression values for all expressed genes. The four main clusters are indicated with colors. (C) Heatmap describing hierarchical clustering of the 28,294 expressed annotated genes using mRNA expression values for all samples. Expression values are scaled per gene so that expression values above the gene average is represented by red, and below average by blue. Eight main clusters have been assigned colors and are denoted a to h. (D) Average expression profiles in tree T1 for each gene expression cluster and distinct sub-clusters (solid white lines). The expression profiles of all individual genes assigned to each cluster is shown as grey lines in the background.*

Results and Discussion

The AspWood resource: high-spatial-resolution expression profiles across cambial growth and wood formation

We have developed the AspWood resource, which contains high-spatial-resolution gene expression profiles across developing phloem and wood forming tissues from four mature, wild-growing aspens (*P. tremula*). The data was

produced using RNA Sequencing (RNA-Seq) of longitudinal tangential cryosections (Uggla et al., 1996; Uggla and Sundberg, 2002) collected from a small block dissected from the trunk of 15-m high, 45-year-old clonal aspen trees during the middle of the growing season (Table S1). The sections (each 15µm thick) were pooled into 25 to 28 samples for each replicate tree. The pooling was based on the estimated tissue composition from anatomical inspection during sectioning (Figure 1A and Table S2), with the aim of sequencing single sections across the cambial meristem, pools of three sections across expanding and SCW forming xylem, and pools of nine sections across maturing xylem. The last sample pool included the previous year's annual ring. Sections that were derived from the mature phloem were pooled into a single sample. We mapped the RNA-Seq reads to the *P. trichocarpa* reference genome (Tuskan et al., 2006; Wullschleger et al., 2013), and classified 29,318 genes as expressed. Of these, 28,294 were previously annotated as protein-coding genes (Table S3). In addition, we identified novel protein coding genes, long intergenic non-coding RNAs (lincRNAs) and other gene fragments with undetermined coding potential, of which 78, 567 and 307, respectively, were expressed (Table S4, S5 and S6). Hence these data expand the *P. trichocarpa* genome annotation with several new transcribed loci. The transcriptome showed high diversity across the entire sample series (Figure S1), with similar numbers of genes observed in all samples, both when considering all expressed annotated genes (range 21,592 to 25,743 genes) or when considering those accounting for 90% of the expression (range 9,181 to 13,420 genes).

We have made the AspWood data available as an interactive web resource (http://aspwood.popgenie.org), enabling visualization and exploration of the data and providing the scientific community with the opportunity to build and test hypotheses of gene function and networks involved in cambial growth and wood formation.

Co-expression network analysis reveals a continuum of transcriptional modules defining cambial growth and wood formation

To provide an overview of the expression dataset, we performed an unsupervised hierarchical clustering analysis of all samples and annotated genes. This identified four major sample clusters (denoted ***i***, ***ii***, ***iii*** and ***iv***) defined by three distinct transcriptome reprograming events (Figure 1AB, Figure S2, Table S2). These three events were assigned developmental context using expression patterns of well-characterized marker genes for phloem differentiation, cambium, radial cell expansion, SCW formation and cell death (see Figure S3 for profiles and documentation), which were further supported by the anatomical data (Figure 1A). The first reprogramming event (***i***/***ii***) occurred between phloem and xylem differentiation, in the middle of the dividing cambial cells. The second event (***ii***/***iii***) marked the end of cell expansion and the onset of SCW formation. The final event (***iii***/***iv***) marked the end of SCW deposition, demonstrating that the late maturation of xylem cells should be considered a defined stage of wood development with a characteristic transcriptome. We have indicated these three transcriptome reprogramming events as reference points in all expression profiles shown in the AspWood resource (e.g. Figure 2).

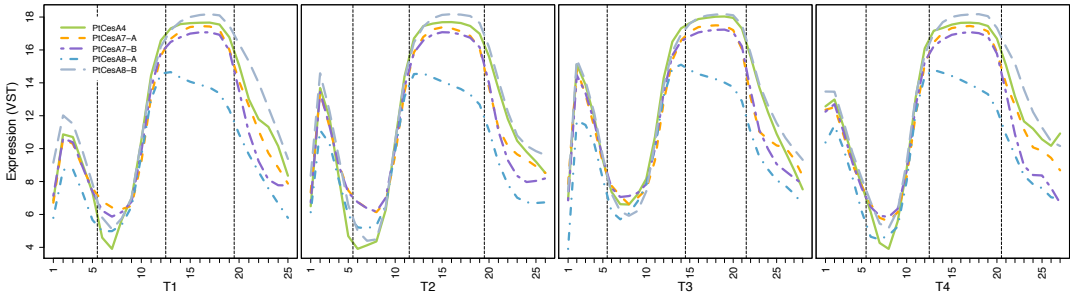

Figure 2. *Expression profiles of the secondary cell wall CesA genes in all four trees T1-T4. CesA4 (Potri.002G257900), CesA7-A (Potri.006G181900), CesA7-B (Potri.018G103900), CesA8-A (Potri.011G069600) and CesA8-B (Potri.004G059600).*

We further used the hierarchical clustering to group genes, dividing the transcriptome into eight major gene expression clusters denoted ***a*** to ***h*** (Figure 1C, Table S7). The expression clusters were highly reproducible in the four replicate trees, demonstrating that the transcriptional landscape of cambial growth and wood formation is under tight genetic control, being robust even under field conditions in a forest setting (Figure 2, Figure S4). With the exception of cluster ***a***, the gene expression clusters, or sub-clusters, showed distinct average expression profiles (Figure 1D). These profiles revealed the major gene expression patterns underlying the three transcriptome reprogramming events identified from the sample clustering.

To develop further insight as to which were the most important genes within the gene expression clusters, and hence also potential regulators, we constructed a co-expression network (see Methods). The network connected pairs of genes with high normalized co-expression (Z-score > 5) and altogether included 14,199 of the 29,246 expressed genes. Genes within each cluster were then ranked according to their centrality in the co-expression network, i.e. the number of co-expressed genes in their network neighborhoods (Table S8). This approach identified several central genes with known roles in cambial growth and wood formation. For example, *SCW cellulose synthase A7* (*PtCesA7-B*) was the most central gene in cluster ***g*** with 165 co-expressed genes. Interestingly, transcription factors (TFs) were more central in the network than other genes (P = 2e-5), and also had more sample-specific expression profiles (i.e. pulse-like profiles with expression restricted to a few consecutive samples, see Methods, P < 2e-16). This most likely reflects the specialized function of many TFs as initiators and coordinators of the transcriptional sub-programs underlying different stages of secondary growth. Several of the *de novo* identified genes were also highly central in the co-expression network. Long intergenic non-coding RNAs central in the co-expression network may represent particularly interesting candidates for functional studies into the importance of this class of transcripts in wood formation (Table S8). However, compared to previously annotated coding genes, a higher proportion of novel genes were not integrated into the network (~30% vs. ~50%), which might reflect their relatively young age (i.e. low conservation, Table S8) (Zhang et al., 2015). Such genes represent potential species-specific adaptations and regulatory mechanisms.

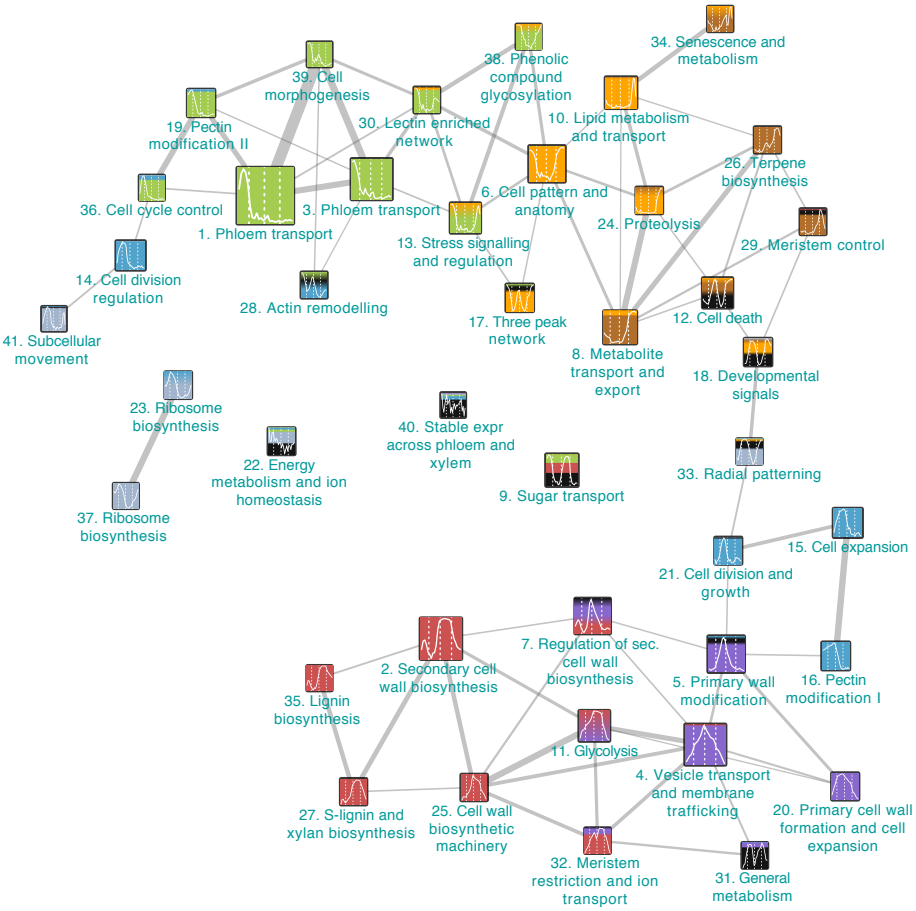

Figure 3. *A clustered version of the co-expression network. Genes with representative expression profiles were identified in the co-expression network (at a Z-score threshold of 5) by iteratively selecting the gene with the highest centrality and a co-expression neighborhood not overlapping with any previously selected genes´ neighborhood. Only annotated protein-coding genes and positively correlated co-expression links were considered (i.e. Pearson correlation > 0). The selected genes and their co-expression neighborhoods (network clusters) were represented as nodes in a clustered network, and given descriptive names (Table S9). The nodes are colored according to the hierarchical clusters in Figure 1, and reflects the proportion of genes in each network cluster belonging to the different hierarchical clusters. The nodes were linked if the neighborhoods overlapped at a Z-score threshold of 4. Link strengths are proportional to the number of common genes. Overlaps of fewer than five genes were not represented by links, and only the 41 network clusters with at least 20 genes were displayed.*

To categorize the transcriptionally regulated biological processes in cambial growth and wood formation, we utilized the co-expression network to identify representative clusters containing genes that were highly co-expressed with central genes in the network (Figure 3). This clustered version of the co-expression network was a result of an unsupervised analysis, relying only on the expression data and not on anatomical annotations or genes with known roles. It thus represented an unbiased description of the central biological processes underlying cambial growth and wood formation. The clustered network contained 41 distinct average expression profiles shared by at least 20 genes. The central

genes in the clusters represented molecular markers of the biological activity while the corresponding clusters represented transcriptional modules underlying important biological processes. We performed Gene Ontology (GO) and Pfam enrichment analyses to assign putative biological functions for these processes (Table S9, Figure 3). The clustered network additionally provided a more detailed view of the transcriptional program responsible for cambial growth and wood formation than the hierarchical clustering, and revealed previously uncharacterized complexity of this process. Noticeably, the different network clusters represented expression peaks that were distributed either monophasically (17 clusters with one peak) or pluriphascially (24 clusters with more than one peak) along the whole developmental gradient, from phloem to the annual ring border. This indicates that anatomical and histochemical markers that are often used to characterize phloem and xylem development are a manifestation of a continuum of transcriptional modules. We next used these network clusters to provide developmental context to the three major transcriptome reprogramming events identified by the hierarchical clustering analysis (Figure 1).

We observed that several network clusters (NC) had expression profiles that each coincided with a single developmental process. NC1 (*Phloem transport*, Figure 3) contained genes that were associated with phloem identity and differentiation, including *ALTERED PHLOEM DEVELOPMENT (APL)* (Bonke et al., 2003), *CLAVATA3/ESR-RELATED 41* (*CLE41)* (Hirakawa et al., 2008; Etchells and Turner, 2010) and *KANADI (KAN)* (Emery et al., 2003; Hirakawa et al., 2008; Etchells and Turner, 2010; Ilegems et al., 2010), and was enriched for genes involved in *sucrose metabolism* (see Table S9 for enrichment details), including *SUCROSE SYNTHASE* (*SUS*) *5* and *6,* which are known to be phloem localized in *A. thaliana* (Barratt et al., 2009). This cluster had a distinct expression peak in the beginning of the sample series, with expression almost completely confined to sample cluster ***i*** (Figure 1), thus marking phloem differentiation and activity. NC14 (*Cell division regulation*, Figure 3) was enriched for the Pfam domain *cyclin*, known to be associated with the progression of the cell cycle, as well as other cell division regulatory genes like *CELL DIVISION CONTROL 2* (*PtCDC2.2*) (Espinosa-Ruiz et al., 2004). This cluster peaked where the hierarchical clustering indicated the first transcriptome reprograming event (***i***/***ii***), thus marking the cambium. NC15 (*Cell expansion*, Figure 3) was enriched for the Pfam domain *pectate lyase*, and included genes such as the alpha-expansin *PtEXPA1* involved in xylem cell expansion (Gray-Mitsumune et al., 2004; Gray-Mitsumune et al., 2008) and the xylem-specific pectate lyase *PtPL1-27* (Biswal et al., 2014). The expression of this cluster was specific to sample cluster ***ii***, thus marking the radial xylem expansion zone. Finally, NC4 (*Vesicle transport and membrane trafficking*, Figure 3) was enriched for genes involved in *vesicle-mediated transport,* and included genes encoding components of the secretory pathway and vesicle transport (e.g. SNARE-like, clathrin and coatomer). This cluster had a broad expression profile, spanning the entire sample series, but with a clear peak where the hierarchical clustering indicated the second transcriptome reprograming event (***ii***/***iii***)**.** This cluster points to the importance of the secretion machinery at this stage of wood formation, and marked where the cell expansion zone ended and the SCW formation zone began (see below). Taken together, these clusters can be used to identify novel candidate regulators of

cambial activity and cambial derivate expansion, in addition to providing novel marker genes associated with the fine-scale expression programs active during secondary growth.

We found that other processes of vascular differentiation were more complex and were associated with several different expression peaks. Examination of both the hierarchical clustering and the network clustering identified clusters with biphasic expression patterns (i.e. profiles with two peaks). Thus this indicated that the same regulatory modules were expressed during, for example, both phloem and a distinct phase of xylem formation, with the same biological processes occurring in these two different cell types. The phloem-xylem biphasic expression pattern was exemplified by NCs 2 (*Secondary cell wall biosynthesis*), 27 (*S-lignin and xylan biosynthesis*) and 35 (*Lignin biosynthesis*) (Figure 3) that included SCW *CesA* genes, xylan genes and lignin biosynthesis genes, respectively. These clusters had expression profiles with a low peak in sample cluster ***i*** and a high and broad peak in sample cluster ***iii***, marking the biosynthesis of SCWs in both xylem cells and phloem fibers. This further illustrated that both xylem and phloem fibers shared similar lignin biosynthetic processes. NC7 (*Regulation of SCW biosynthesis*, Figure 3) contained the two master regulatory switches for SCW formation; *MYB46* ortholog *PtMYB021* and two *MYB83* orthologs (McCarthy et al., 2010). This cluster also included several laccases orthologous to LAC2, LAC5 and LAC17, as well as dirigent like proteins and cytochrome P450 encoding genes, and was enriched for *oxidation-reduction process*, possibly associated with the lignification initiated at the end of the cell expansion phase. The cluster was characterized by a sharp peak marking the initiation of SCW formation in the xylem and a lower sharp peak on the phloem side that may coincide with phloem fiber formation. This expression pattern is consistent with this cluster including regulatory switches that induce the broader expression profiles of the structural genes in clusters such as NC 2, 25, 27 and 35, and is an example of the finding that TFs had significantly narrower expression domains than other genes. Finally, NC34 (*Senescence and metabolism*, Figure 3) was enriched for the Pfam domain *late embryogenesis abundant protein*, and included an ortholog of *BIFUNCTIONAL NUCLEASE I* (*BFN1*), which is implicated in the post-mortem destruction of dead xylem vessel nuclei (Ito and Fukuda, 2002). The cluster had two sharp peaks, with the first peak appearing late in sample cluster ***iii***, before the third and final transcriptome reprograming event (***iii***/***iv***), and the second peak occurring late in sample cluster ***iv***. The decrease in expression of these two peaks coincides with the loss of viability in vessels elements and fibers, respectively (Table S2), thus marking the border of living vessel elements and fibers.

Taken together, the clustered network provided an unbiased framework enabling placement of biological processes during cambial growth and wood formation, and revealed a greater complexity than had been realized on the basis of previous, lower spatial and transcriptome resolution. While many of the network clusters could be associated with known process, several remained poorly characterized and warrant future attention. A notable feature of many of the network clusters, and in particular the uncharacterized ones, was the presence of two or three peaks of expression, possibly representing biological processes active in different cell types or at different stages during the differentiation of a specific cell type.

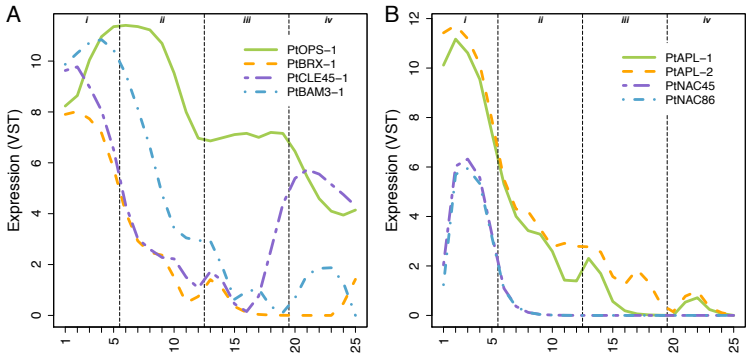

Figure 4. *Conserved expression patterns for regulators of phloem development between primary and secondary growth. (A) Expression profiles for PtOPS, PtBRX, PtCLE45 and PtBAM3. PtOPS genes, here represented by PtOPS-1, were highly expressed in cambium while PtBRX, PtCLE45 and PtBAM3 were highly expressed in the secondary phloem. (B) Expression profiles for PtAPL and PtNAC45/87. PtAPL and PtNAC genes were highly expressed in the secondary phloem. For the complete list of genes, see Table S10.*

Similarities in the regulation of primary and secondary phloem development

The two plasma membrane-associated proteins OCTOPUS (OPS) and BREVIS RADIX (BRX) are expressed in procambium/sieve tube elements and sieve tube element precursor cells, respectively in the root of *A. thaliana*, and are required for the correct specification of these cells during primary vascular development (Scacchi et al., 2010; Truernit et al., 2012). Loss-of-function mutants display similar phenotypes for both genes, where individual protophloem cells cannot differentiate and create gaps in protophloem strands. In contrast, in the sieve tube element precursors, the CLAVATA3/EMBRYO SURROUNDING REGION 45 (CLE45) peptide inhibits the early protophloem development by signaling through the LRR-RLK receptor BARELY ANY MERISTEM3 (BAM3) (Depuydt et al., 2013; Rodriguez-Villalon et al., 2014). In agreement with these functions, we identified that orthologs of *OPS* displayed high expression across the cambium, whereas *PtBRX, PtCLE45* and *PtBAM3* genes displayed high expression in the differentiating phloem (sample cluster ***i***, Figure 4A). Another key gene in regulating phloem development is encoding the MYB transcription factor ALTERED PHLOEM DEVELOPMENT (APL) (Bonke et al., 2003). APL acts as a positive regulator of phloem identity in *A. thaliana*, and *apl* plants show ectopic xylem formation where phloem cells normally form (Bonke et al., 2003). *PtAPL* and orthologs of *APL* targets *NAC45/86* also show high expression in the differentiating phloem in aspen (Figure 4B) (Furuta et al., 2014). This suggests that components regulating primary phloem development in *A. thaliana* roots may function similarly during secondary phloem development in tree stems. Moreover, we used AspWood to identify several genes co-expressed with these known regulators, including several TFs from a broad range of families (Table

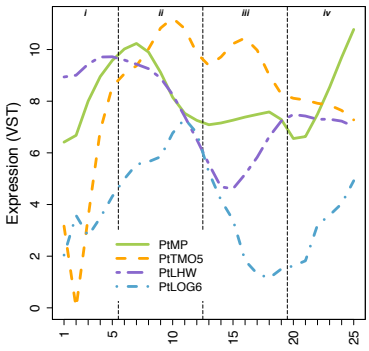

Figure 5. *Regulators of periclinal cell divisions. Expression profiles for PtMP, PtTMO5, PtLHW and PtLOG6. High expression levels for all genes overlap in the expanding xylem. For the complete list of genes, see Table S10.*

S10). These co-expressed genes represent new candidates potentially related to phloem development in aspen.

Regulation of periclinal cell divisions in cambium

In *A. thaliana,* establishment of the vascular tissues in the embryo, and the periclinal cell division activity of the root procambium, is regulated through the basic helix-loop-helix (bHLH) transcription factors TARGET OF MONOPTEROS 5 (TMO5) and LONESOME HIGHWAY (LHW) acting downstream of auxin signaling (Schlereth et al., 2010; De Rybel et al., 2013). TMO5 and LHW function as hetero-dimers, and co-accumulation of their transcripts in the xylem precursor cells of the root meristem is necessary and sufficient to trigger periclinal cell divisions of the procambium cells (De Rybel et al., 2013). This non-cell autonomous stimulation of cell division activity appears to be through local biosynthesis of cytokinin by the LONELY GUY3 (LOG3) and LOG4 proteins (De Rybel et al., 2014; Ohashi-Ito et al., 2014). In return, cytokinin acts as a mobile signal to promote periclinal cell divisions in the procambium.

We found that although *PtTMO5* and *PtLHW* genes had complex expression patterns in the wood-forming zone of aspen, high levels of both genes intersected in the xylem expansion zone (sample cluster ***ii***, Figure 5). Likewise, the *Populus* ortholog of *LOG4* displayed high expression in this region. Genes similar to *MP/ARF5,* encoding an upstream regulator of the TMO5-LHW dimer (Schlereth et al. 2010), were expressed both in the cambium and in the expanding xylem. Taken together the expression profiles of orthologous *Populus* genes suggest that a bHLH complex involving similar components to *A. thaliana* roots might regulate periclinal cell divisions during cambial growth in trees. Moreover, as in *A. thaliana* roots, expression of *LOG* suggests a movement of cytokinins to stimulate cell divisions, in this case from the expanding xylem to cambial cells.

Primary cell-wall-polysaccharide biosynthetic genes continue to be expressed during secondary cell wall deposition in xylem tissues

In *Populus* xylem, primary and secondary cell wall layers have very different polysaccharide composition (Mellerowicz and Gorshkova, 2012). The primary wall layers are abundant in pectins, such as homogalacturonan (HG) and

rhamnogalacturonan I (RG-I), and they contain hemicelluloses such as xyloglucan, arabinoglucuronoxylan and mannan. In contrast, the SCW layers are rich in glucuronoxylan, and contain only small amounts of mannan. Cellulose is an important component in both layers, but the proportion of cellulose is significantly higher in SCW layers. These differences indicate that the cell wall polysaccharide biosynthetic machinery undergoes significant rearrangement during primary-to-secondary wall transition. The spatial resolution of AspWood enabled us to characterize the transcriptional nature of this shift.

We identified the closest *Populus* orthologs of *A. thaliana* genes encoding glycosyl transferases involved in the biosynthesis of HG (Atmodjo et al., 2011), RG-I (Harholt et al., 2006; Liwanag et al., 2012), RG-II (Egelund et al., 2006), and xyloglucan (Cavalier and Keegstra, 2006; Zabotina et al., 2008; Chou et al., 2012) (Table S10). As expected, the majority of these genes were highly expressed in primary walled cambial and radially expanding tissues (sample cluster ***ii***, Figure 6A). Interestingly, although the expression peaked before the onset of SCW deposition, a majority were also significantly expressed during SCW deposition (sample cluster ***iii***). This suggests that there may be continuous biosynthesis of pectin and xyloglucan during SCW biosynthesis. The incorporation of these polymers could still occur in the primary wall layer, as suggested for xyloglucan in developing aspen fibers based on immunolabelling localization patterns (Bourquin et al., 2002), and in the parenchyma cells. Moreover, transcripts of many enzymes involved in pectin and xyloglucan metabolism, such as pectate lyases, and xyloglucan endotransglycosylases (XETs), are also abundant in the SCW formation zone (Mellerowicz and Sundberg, 2008), and XET activity is localized in secondary wall forming fibers (Bourquin et al., 2002) indicating a continuous metabolism of these polysaccharides. Interestingly, the pectin biosynthetic genes *PtGAUT1-A, PtARAD1-A, PtGALS2-A* and *PtGALS2-B* had a distinct peak in mature xylem (sample cluster ***iv***, Figure 6A). These peaks may correspond to the biosynthesis of protective and isotropic layers deposited as tertiary wall layers after the deposition of SCW in the contact and isolation ray cells, respectively (Fujii et al., 1981; Murakami et al., 1999). These late changes of pectin structure during SCW formation may also be associated with the pectin modification of the bordered pits during the maturation of vessels (Herbette et al., 2015).

A. thaliana SCW cellulose synthase complexes contain equimolar trimers of CesA4, 7 and 8 (Hill et al., 2014), and *Populus* orthologs forming similar complexes have been identified (Song et al., 2010). In AspWood, the corresponding transcripts showed a very specific expression pattern, marking the onset and end of SCW biosynthesis in the xylem (sample cluster ***iii***), as well as fiber differentiation in phloem (sample cluster ***i***, Figure 6B). The other *CesA* genes, orthologous to *A. thaliana* "primary wall *CesAs*" (Kumar et al., 2009), showed varied expression patterns, but the majority were highly expressed in the cambial and radially expanding tissues (Figure 6B), consistent with their role in cellulose biosynthesis in primary walls. Interestingly, there were also primary wall *CesAs* that showed similar expression patterns as pectin and xyloglucan biosynthesis genes, with high transcript abundances during SCW biosynthesis bringing into question their primary wall specificity. A functional role for "primary wall CesAs" during SCW

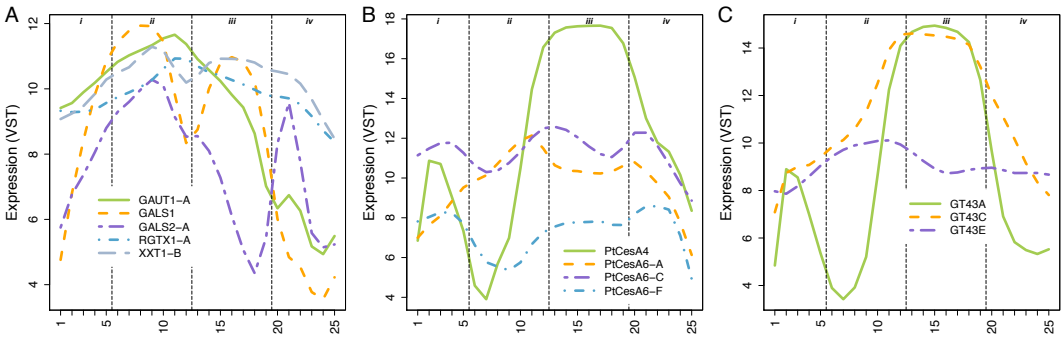

Figure 6. *(A) Expression profiles for pectin and xyloglucan biosynthetic genes. All genes are highly expressed during primary wall biosynthesis, but also at later stages during xylem development. i) representative expression for homogalacturonan biosynthesis genes (illustrated by PtGAUT1-A), ii) expression pattern of PtGALS1, iii) representative pattern of RG-I biosynthesis genes (represented by PtGALS2-A), and iv) representative expression pattern for three of the putative xyloglucan biosynthesis genes (illustrated by PtXXT1-B). For a list of identified putative pectin and xyloglucan biosynthesis genes see Table S10. (B) Expression patterns for CesA genes. i) Members responsible for cellulose biosynthesis in the secondary wall layers are all induced in SCW biosynthesis zones in the xylem, and phloem (illustrated by PtCesA4) ii) Members classified as primary wall CesAs typically peak in primary wall biosynthesis zone, but are also highly expressed during later stages of xylem differentiation, (illustrated by PtCesA6-A), iii-iv) and some members even peak during these later stages (illustrated by PtCesA3-C, and 6-F). For a complete list of putative CesA genes see Table S10. (C) The GT43 gene family responsible for xylan biosynthesis comprise three clades, A/B, C/D and E, each having different expression here illustrated by i) PtGT43A, ii) PtGT43C and iii) PtGT43E. Expression profiles support the hypothesis that PtGT43A/B and PtGT43C/D are members of secondary wall xylan synthase complex, whereas PtGT43E and PtGT43C/D are members of the primary wall xylan synthase complex. For a complete list of genes co-regulated with the three clades of GT43 genes, see Table S10.*

biosynthesis is supported by the presence of protein complexes containing the corresponding CesAs proteins in xylem cells synthesizing SCW layers (Song et al., 2010). Thus, the "primary wall *CesA* genes" might have a more general role than previously thought, including the possibility that they enter into complexes with SCW CesAs (Carroll et al., 2012).

Arabinoglucuronoxylan of primary cell wall layers and glucuronoxylan of SCW layers have the same backbone of β-1,4-Xyl*p* residues, which is thought to be synthesized by a heteromeric xylan synthase complex in the Golgi (Oikawa et al., 2013). At least three interacting glycosyltranferases (GT) of this complex have been identified, i.e. one GT47 member (IRX10/10L) and two GT43 members (IRX9/9L and IRX14/14L) (Jensen et al., 2014; Urbanowicz et al., 2014; Mortimer et al., 2015; Zeng et al., 2016). In *Populus*, IRX9/9L function is performed by *Pt*GT43A, B, or E, and IRX14/14L function by *Pt*GT43C or D (Lee et al., 2011; Ratke et al., 2015). In AspWood, expression of *PtGT43A* and *B* was almost identical and closely resembled that of SCW CesAs, whereas expression of the genes encoding their interacting partners *Pt*GT43C or D was slightly different, with higher transcript levels in primary-walled tissues (Figure 6C). *PtGT43E* showed a very

different expression, with a broad peak in primary walled xylem cells. These observations support the recently suggested concept that separate primary and secondary xylan synthase complexes exist, similar to primary and secondary cellulose synthase complexes (Mortimer et al., 2015; Ratke et al., 2015), with *Pt*GT43E (IRX9L) and *Pt*GT43C/D (IRX14) forming the primary xylan synthase complex, and *Pt*GT43A/B (IRX9) and *Pt*GT43C/D (IRX14) forming the secondary xylan synthase complex. Distinction between expression patterns of *Pt*GT43A/B and *Pt*GT43C/D in AspWood highlighted the high resolution of the data and demonstrated the potential for discovering new genes involved in wood formation.

Taken together, the expression profiles of the pectin, xyloglucan and cellulose primary wall biosynthetic genes suggest that the primary wall biosynthetic program is not terminated at the onset of secondary wall biosynthesis. It rather continues in the background of the more spatially confined expression of SCW genes such as CesA 4/7/8 and *Pt*GT43A/B. Indeed, when SCW deposition is terminated in long-lived cell types such as parenchyma cells or gelatinous-fibers in tension wood, a subsequently deposited layer containing pectin and xyloglucan can be observed in these cells that resembles the primary cell wall layer (Fujii et al., 1981; Murakami et al., 1999; Mellerowicz and Gorshkova, 2012). Our analyses additionally revealed that some genes of the SCW biosynthesis program are also active during the period of primary cell wall formation. This was exemplified by *Pt*GT43C/D and its co-expression network neighborhood, which included many proteins involved in general cellular functions associated with cell wall biosynthesis, such as vesicle trafficking, transport, sugar nucleotide metabolism and general cell wall acetylation machinery (Table S10).

Spatially separated expression of phenoloxidases may enable site and cell type specific lignification

Cell wall lignification results from the secretion of differently methoxylated 4-hydroxyphenylpropanoids, called monolignols, which are radically oxidized by cell wall resident phenoloxidases (laccases and peroxidases) to cross-couple the monolignols into an insoluble polyphenolic polymer in the cell wall (Boerjan et al., 2003; Barros et al., 2015). The high spatial sampling in the present dataset enabled us to resolve the expression domain of lignification during the formation of the different secondary xylem cell types.

Genes associated to lignin formation include two groups: (i) monolignol biosynthesis-coding genes, including 11 enzymatic steps ranging from phenylalanine ammonia lyases (PALs) to cinnamyl alcohol dehydrogenases (CADs), and (ii) phenoloxidases, including laccases and peroxidases (Boerjan et al., 2003; Barros et al., 2015). In *Populus*, 92 genes encoding monolignol biosynthetic genes have been identified (Shi et al., 2010) (Table S10). We found that not all the paralogs of each family were expressed during xylem formation (Fig. 7A, Table S10), suggesting that these genes may be involved in the biosynthesis of phenolic compounds other than wood lignin. The expressed monolignol biosynthetic genes generally exhibited a bi-phasic expression profile, with one peak in the differentiating phloem (i.e. sample cluster ***i***) and one peak in

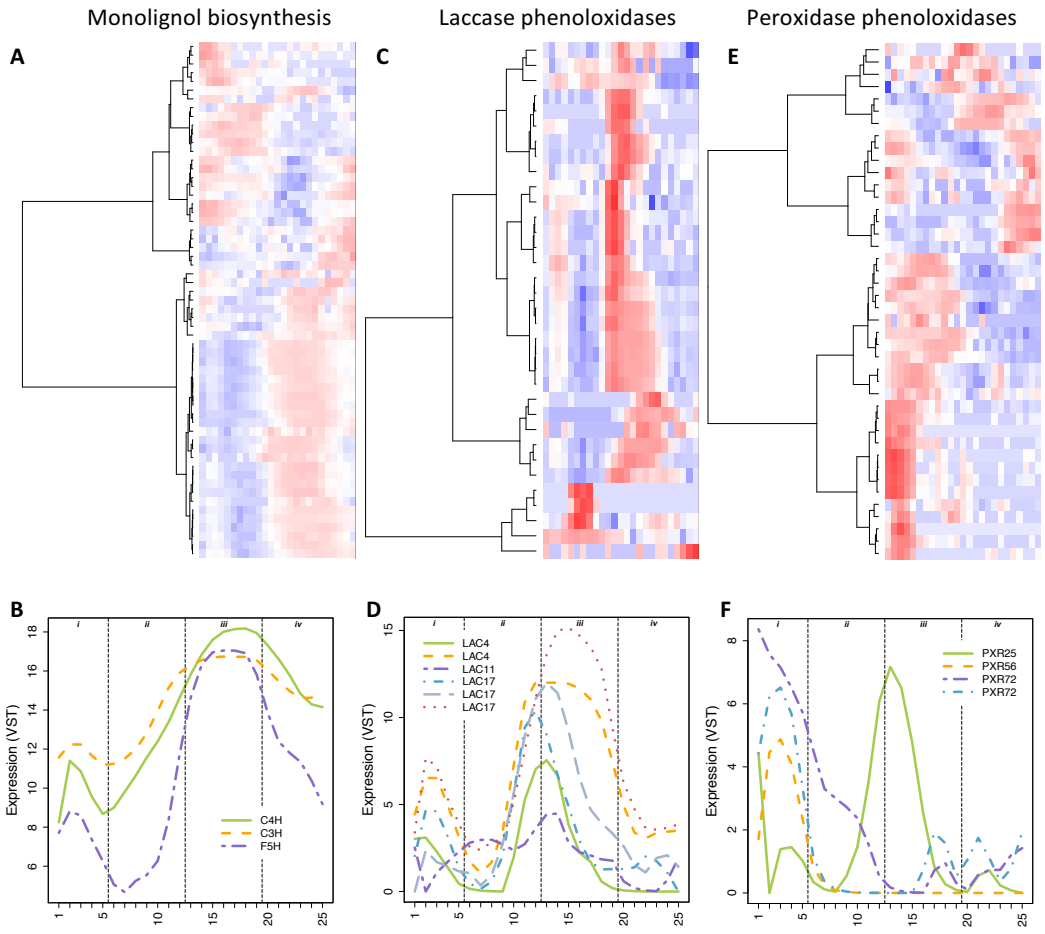

Figure 7. *Expression profiles of genes involved in the lignification of xylem cells. (A) Hierarchal clustering and heatmap of the 59 monolignol biosynthesis genes expressed in AspWood. Expression values are scaled per gene so that expression values above the gene average is represented by red, and below average by blue. (B) Expression profiles of C4H (Potri.013G157900), C3H (Potri.006G033300) and F5H (Potri.005G117500). (C) Hierarchical clustering of the 34 expressed laccase phenoloxidases, and (D) expression profiles of representative genes: LAC4 (Potri.001G248700, Potri.016G112100), LAC11 (Potri.004G156400) and LAC17 (Potri.001G184300, Potri.001G401300 and Potri.006G087100). (E) Hierarchical clustering of the 42 expressed peroxidase phenoloxidases, and (F) expression profiles of representative genes: PXR25 (Potri.006G069600), PXR56 (Potri.007G122200) and PXR72 (Potri.005G118700 and Potri.007G019300).*

the maturing xylem with high expression even after vessel and fiber death (i.e. sample clusters ***iii*** and ***iv***, Fig. 7AB). However, some paralogs were expressed only in the phloem or in the late maturing xylem (Fig. 7A).

In *Populus*, 165 genes encoding phenoloxidases have been identified (Lu et al., 2013) (Table S10). Both peroxidases and laccases have been shown to be active in wood forming tissues of *Populus* and to be capable of polymerizing both mono-methoxylated (guaiacyl G) and bi-methoxylated (syringyl S) monolignols into lignin-like polymers *in vitro* (Christensen et al., 1998; Ranocha et al., 1999; Sasaki et al., 2004; Sasaki et al., 2008). More recently, laccases *LAC4*/*IRX12*, *LAC11* and

LAC17 were demonstrated to act synergistically during vessel element and fiber lignification in *A. thaliana* (Zhao et al., 2013) and in *Populus* (Lu et al., 2013). Lignin formation has also been shown to be dependent on the additive activities of multiple peroxidases, including *PXR2*, *PXR25*, *PXR71* and *PXR72* in *A. thaliana* xylem (Herrero et al., 2013; Shigeto et al., 2015) or *PXR3* (Potri.003G214800) in *Populus* (Li et al., 2003). We found that many of these phenoloxidases were expressed in aspen xylem (35 out 56 for laccases and 42 out of 109 for peroxidases, Table S10). In contrast to the very broad expression pattern exhibited by monolignol biosynthetic genes, laccases and peroxidases expressed in the xylem showed very narrow expression patterns. Most laccase clusters peak at different phases of SCW formation (sample cluster ***iii***, Fig. 7CD). Comparison of the expression profiles of these genes with the cell viability measurements (Table S2) suggested that laccases show specific expression for vessel and/or fibers. Few peroxidases were coexpressed with laccases, and most of them were expressed in the phloem and/or the cambium (i.e. sample clusters ***i*** and ***ii***) and in the late maturing xylem (i.e. sample cluster ***iv***, Fig. 7EF). This suggests that phloem fiber lignification is mostly dependent on peroxidases. Thus, a clear spatial separation between laccases and peroxidases during secondary xylem formation suggests that in aspen these enzymes are not acting redundantly, as previously shown in *A. thaliana* (Zhao et al., 2013).

Taken together, the transcriptional regulation of lignification associated genes during *Populus* xylem formation suggests that monolignol biosynthesis is not the primary determinant controlling the timing, localization and composition of lignin in *Populus* xylem (Fig. 7A). In contrast, phenoloxidases exhibited more specific expression profiles with laccases associated with vessel element and fiber lignification (Fig. 7C) and peroxidases generally associated with phloem, cambium and mature xylem (Fig. 7E). Thus, phenoloxidases are acting both synergistically and additively to enable the lignification of xylem in specific cell-types.

Transcriptional regulators of wood development

The differentiation of xylem vessel elements and fibers is regulated by NAC domain transcription factors (for reviews, see Zhong and Ye (2014), Ruzicka et al. (2015) and Ye and Zhong (2015)). In *A. thaliana*, expression of *VASCULAR-RELATED NAC-DOMAIN 6* (*VND6*) and *VND7* specifies vessel element cell fate (Kubo et al., 2005). In addition, VND1 to 5 also induce the transcription of SCW genes and vessel element formation (Zhou et al., 2014). Xylem fiber differentiation is mediated by another class of NAC transcription factors, namely members of the *SECONDARY WALL-ASSOCIATED NAC DOMAIN* 1 (*SND1*) and *NAC SECONDARY WALL THICKENING PROMOTING* 1 (*NST1*) clade (Zhong et al., 2006). The NAC domain transcription factors SND2 and 3 are thought to serve a similar role in *A. thaliana*.

The *P. trichocarpa SND*/*VND* orthologs separate into five distinct phylogenetic clades (Figure 8A). Within a given clade, paralogs were typically highly co-expressed with one notable exception; members of the *VND3* clade displayed similar expression patterns to either the *VND6* or *SND2* paralogs (Figure 8A-E). For members of the *VND6* clade, expression peaked either at the zone of cell

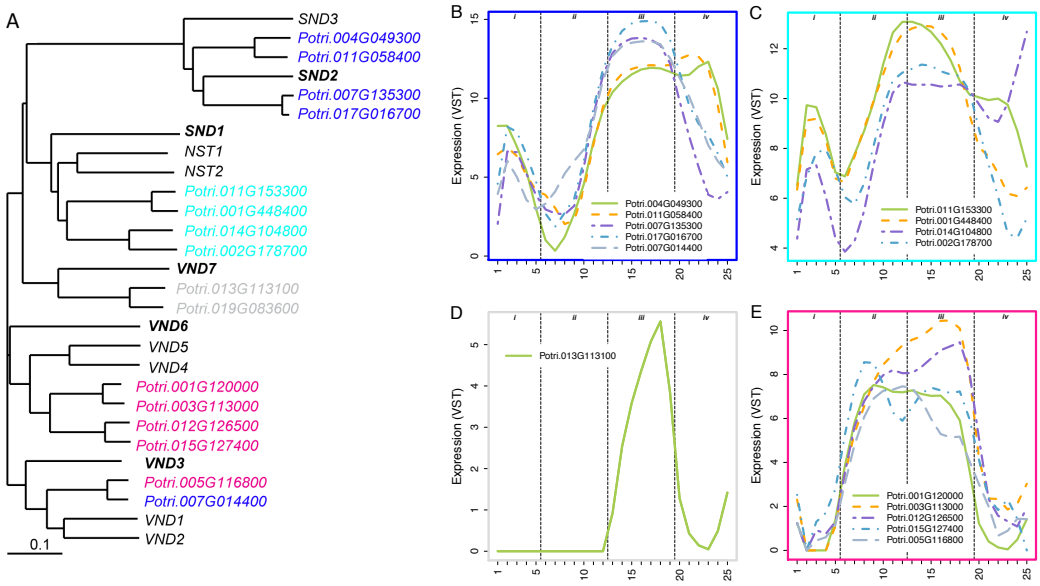

Figure 8. *NAC domain transcription factors are expressed in distinct patterns corresponding to phylogenetic clustering. (A) Phylogenetic tree of Populus wood associated NAC-domain transcription factors with their closest orthologs in A. thaliana. Colors indicate co-expression. (B) Expression profiles of Populus SND2 (Potri.007G135300; Potri.017G016700) and SND3 (Potri.004G049300; Potri.011G058400) orthologs compared to a VND3 ortholog (Potri.007G014400). (C) Expression profiles of Populus SND1 orthologs (Potri.001G448400; Potri.002G178700; Potri.011G153300; Potri.014G104800). (D) Expression profile of a Populus VND7 ortholog (Potri.013G113100). (E) Expression profiles of Populus VND6 orthologs (Potri.001G120000; Potri.003G113000; Potri.012G126500; Potri.015G127400).*

expansion (sample cluster ***ii***) or further inwards at the end of the SCW zone (sample cluster ***iii***, Figure 8E). Early expression of *VND6* paralogs is in line with a role as a vessel element identity gene, with cell fate being specified at the onset of radial expansion. The continuous expression of VND6 during SCW deposition, however, points to an additional role of *VND6* paralogs other than cell type specification. Indeed, this extended expression of *VND6* during SCW formation has also been observed in *A. thaliana* xylem vessel forming cell cultures (Derbyshire et al., 2015). While one of the *VND7* paralogs, which in *A. thaliana* specifies protoxylem, was not expressed in the developing xylem, expression of the second *VND7* paralog increased sharply in the SCW deposition zone (sample cluster ***iii***, Figure 8D), overlapping with the inner peak of the *VND6* paralogs. In the extended co-expression network neighborhood (77 genes at a z-score threshold of threshold 3), *Populus VND7* positively correlated with *METACASPASE 9* (*MC9*) and was negatively linked to *SUCROSE SYNTHASE 3 (SUS3)* expression. While *MC9* participates in the autolysis of xylem vessel elements (Bollhoner et al., 2013), *SUS3* may be involved in making carbon available for cell wall biosynthesis (Gerber et al., 2014). Together, the neighborhood may suggest that maximal expression of the *Populus VND7* ortholog and the later peak in the expression pattern of the *VND6* paralogs mark the transition from the SCW to the maturation phase in xylem vessel elements. Thus, the spatial resolution of our sample series enabled us to

define the two clearly separated expression domains of genes active in these processes.

Compared to the *VND6* paralogs (Figure 8E), expression of the *SND1* paralog was shifted slightly towards maturing xylem, but reached maximal values before the transition to the SCW formation zone (i.e. sample cluster ***iii***, Figure 8C). However, an almost perfect overlap with the SCW formation marker *CESA8* was found for the expression patterns of the *SND2* paralogs (Figure 8B). Unlike the *VND6* paralogs, both *SND1* and *SND2* paralog abundance increased gradually from the cambium towards the phloem in sample cluster ***i***, suggesting a role in SCW formation in phloem fibers. This suggests that the role of *SND1* and *SND2* paralogs is specific to fibers but does not discriminate between phloem and xylem. Taken together, gene expression of different NAC domain transcription factors involved in xylogenesis was induced in four distinct, non-overlapping domains. Early onset of expression of both *VND6* and *SND1* is in support of their role in the specification of vessel elements and fibers. However, their expression remained high during the entire SCW phase, indicating a role more directly connected to the deposition of wall polymers and/or *post-mortem* autolysis.

To investigate the power of the co-expression network in identifying transcription factor targets, we analyzed the co-expression neighborhood of SND1. In *Populus,* 76 direct target genes of SND1 were recently identified using transactivation assays and chromatin immunoprecipitation (ChIP) in protoplast-derived xylary tissue with a low degree of differentiation (Lin et al., 2013). Another 34 genes (hereafter, non-ChIP targets) were identified as transactivation targets of SND1 and its paralogs based on different lines of experimental evidence (Ye and Zhong, 2015) (see Table S10 for all target genes). These 110 putative direct targets of SND1 mapped to 76 annotated protein coding genes with expression in our wood series. Sixty-eight (89%) formed a connected co-expression network with the four *SND1* paralogs across the differentiating wood using a co-expression threshold of 3 (Figure S5). Interestingly, the majority of the ChIP targets were negatively correlated with the *SND1* paralogs (sub-cluster II and III in Figure S5), indicating that SND1-like transcription factors likely suppress transcription of these targets.

A majority of paralogs show differential expression during wood formation

In *P. trichocarpa*, a whole genome duplication (WGD) occurred relatively recently (58 million years ago, (Dai et al., 2014). WGDs are believed to be major contributors to the evolution of novel function in genomes (reviewed in Hermansen et al. (2016)), and eventually to species diversification. Gene duplicates resulting from WGDs, hereafter called paralogs, may evolve through different processes; a) sub-functionalization, where the genes retain different parts of the ancestral function, b) neo-functionalization, where one gene retains the ancestral function while the other evolves a new function, or c) non-functionalization, where one of the genes is eliminated by random mutations. These processes may act both on protein function and on gene expression, with regulatory divergence being particularly important for evolution of plant development (Rosin and Kramer, 2009). A previous microarray study in *Populus* covering 14 different tissues indicated that nearly half of the paralog pairs had

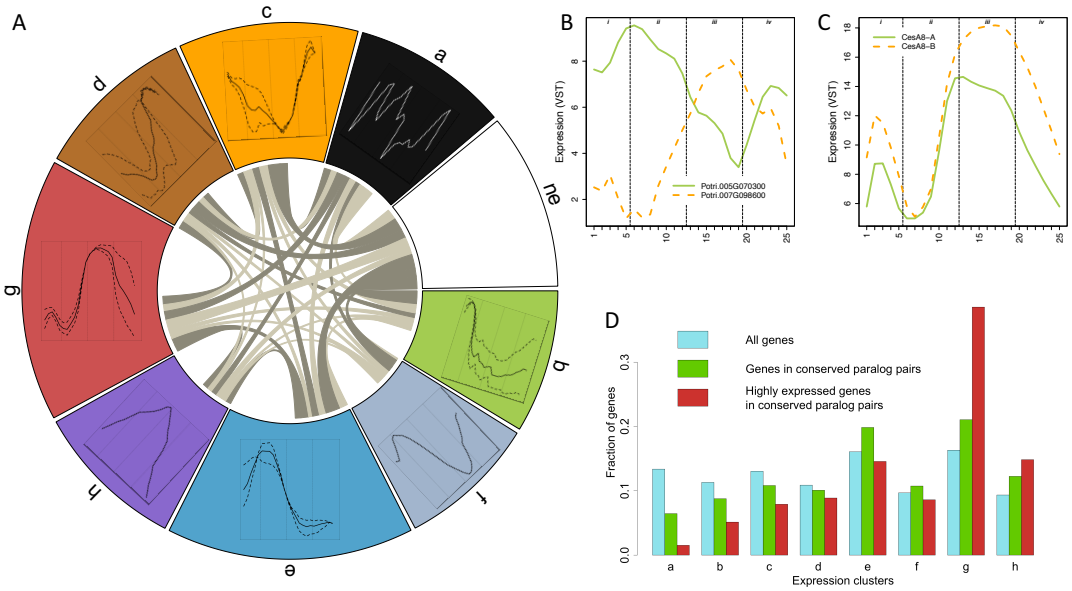

Figure 9. *WGD paralog expression. (A) Circus plot of regulatory diverged and conserved paralogs. The clusters are ordered according to their peak of expression along the wood developmental gradient. An additional cluster for genes not expressed (ne) in our data is also added. Each cluster occupies a share of the circle´s circumference proportional to the number of genes in that cluster belonging to a paralog pair. Paralogs expressed in two different clusters are shown by links. The width of a link is proportional to the number of paralogs shared between these clusters (i.e. diverged pairs). Only pairs in different clusters and with an expression Pearson correlation < 0.5 were considered diverged. Links representing more pairs than expected by chance (P < 0.0001) are colored in a darker tone. The share of a cluster without links represents the proportion of paralogs where both genes belong to that same cluster (i.e. conserved pairs). (B) Example of paralogs with highly diverged profiles (correlation = -0.78): FAAH (fatty acid amide hydrolase, Potri.005G070300 (cluster **e**) and Potri.007G098600 (cluster **g**)). (C) Previously published real-time PCR data of CesA genes showed that PtCesA8-B (Potri.004G059600) was expressed higher than PtCesA8-A (Potri.011G069600) in secondary phloem and xylem (Takata and Taniguchi, 2015). Although these genes have highly similar expression profiles (Pearson correlation = 0.93), and are therefore considered as conserved according to our definition (both belong to cluster **g**), the data confirm the difference in expression levels and show that even more subtle regulatory divergence can be detected in the dataset. (D) The distribution of different gene sets among expression clusters: (blue) all 28,294 expressed, annotated genes in our data, (green) all 3,729 paralog pairs with conserved expression (i.e. in the same expression cluster), (red) 721 genes with high expression (among the top 5% most highly expressed genes in our data; 1,413 genes) and in paralog pairs with conserved expression.*

diverged in their expression (Rodgers-Melnick et al., 2012). We used our RNA-seq data to map the regulatory fate of WGD paralogs in cambial growth and wood formation. Of the 9,728 paralogous gene pairs identified in the *P. trichocarpa* genome by sequence similarity and synteny (see Methods), 8,844 had at least one gene expressed in our data. Of these paralogs, 3,185 (36%) were regulatory divergent, as defined by their presence in different expression clusters *and* by their expression correlation being below 0.5 (Figure 9A, Table S11A). In addition, another 1,930 paralogs (22%) had only one gene expressed in our data. These may

represent cases of non-functionalization, or cases where one gene is expressed in another tissue or condition. Paralogs with diverged expression were often found to be enriched in cluster-pairs with relatively similar average expression profiles (e.g. present in clusters ***g*** and ***d***) (Figure 9A, Table S11B), but paralogs with more diverged expression were also observed. For example, 1,163 paralogs (17%) had negatively correlated expression profiles, of which 156 displayed mirrored profiles (correlation < -0.50, see e.g. Figure 9B). Our estimate of regulatory divergence can be considered conservative, representing cases where paralogs display distinctly different expression profiles across cambial growth and wood formation. However, the high spatial resolution of our dataset also enabled the identification of far more subtle differences, such as the previously reported differences in expression levels between *CesA8-A* and *CesA8-B* in secondary phloem and xylem (both located in cluster ***g***, Figure 9C (Takata and Taniguchi, 2015). While the processes of sub-, neo- or non-functionalization drives the divergence of paralog pairs, the gene dosage balance hypothesis may explain why some paralog pairs have retained similar expression profiles (Yoo et al., 2014). Interestingly, a high proportion of the paralog pairs with similar profiles were found among the 5% most abundant transcripts in our dataset (51%, p = 5e-92). In particular cluster ***g***, including genes expressed during SCW formation (sample cluster ***iii***), contained a substantially higher fraction of such highly expressed and regulatory conserved paralogs than other clusters (Figure 9D). This suggests that it has been advantageous for *P. tremula* to maintain higher levels of certain genes involved in SCW formation than could be achieved by a single ancestral gene copy. A second copy can also make the system more robust to perturbations.

Conclusions

Using the AspWood high spatial resolution gene expression resource, we revealed previously uncharacterized complexity of cambial growth and wood formation, with numerous well-defined co-expression clusters being observed continuously from phloem to the previous year's annual ring. Moreover, the high spatial resolution of the data enabled the discovery of unexpected expression domains of known genes, for example genes involved in ligno-cellulose biosynthesis and master regulators of SCW, in addition to spatially resolving the expression domains of genes involved in lignification. The data also enabled us to analyze the evolutionary fate of highly sequence-similar paralogs arising from a recent whole genome duplication, for which we show that a majority of paralogs displayed diverged expression profiles across wood forming tissues.

The AspWood web resource provides the scientific community with an interactive tool for exploring the co-expression network of cambial growth and wood formation in a model angiosperm forest tree. For example, we showed how the network can be used to identify the targets of SND1, with a majority of experimentally obtained targets being co-expressed with this regulator. This analysis additionally revealed that several of these targets were negatively correlated with SND1 expression, indicating a novel role for this regulator as a transcriptional repressor.

Methods

Sampling and RNA extraction

Wood blocks were collected from four independent naturally growing *P. tremula* clones (tree IDs T1, T2, T3 and T4) from Vindeln, north Sweden. Hand sections were taken from freshly collected stem material for analysis of cell viability using nitrobluetetrazolium (Courtois-Moreau et al., 2009). 15 micron thick longitudinal sections were cut using a cryo-microtome (Uggla et al., 1996) and stored at -80C. Cross-sections were taken during the process of sectioning and imaged with a light microscope (Lieca). These images were used to characterize the different tissue types present. All cryosections were deemed to originate from one of four developmental zones: phloem, cambium, early/developing xylem and mature xylem. No obvious biotic or abiotic stresses were evident at the time of sampling and anatomically discernible developing tension wood was avoided. The total number of sections varied from 105 to 135 in the four different trees, and covered the entire current year's growth.

Individual sections were thawed in QIAzol Lysis Reagent (Qiagen, Manchester, UK) and homozenised using a Retsch mixer mill. Total RNA and small RNA were extracted from section pools as indicated in Table S2. The miRNeasy Mini Kit (Qiagen, Manchester, UK) was used for extractions. For Total RNA, one elution of 35 µl was made. For small RNA, two elutions of 25 µl each were made and pooled.

Total RNA was quantified using a Nanodrop 1000 (Thermo Scientific, Wilmington, USA) and quality was assessed using the Agilent 2100 Bioanalyzer ((Agilent Technologies, Santa Clara, USA) using Pico chips (Agilent Technologies, Santa Clara, USA). A 100 ng aliquot was used for amplification. If the volume of the 100 ng aliquot was larger than 5 µl, it was dried using a speed-vac. In some cases, less than 100 ng total RNA was available, in which case the maximum possible amount was used. Total RNA was amplified using Ambion MessageAmp™ Premier RNA Amplification Kit (Ambion, Thermo Scientific, Wilmington, USA following manufacturer's instructions. Using 100 ng of total RNA as starting material, we obtained 10-63 µg of amplified RNA (aRNA). aRNA was quantified using a Nanodrop and integrity checked using a Bioanalyzer with Nano chips (Agilent Technologies, Santa Clara, USA). 2.5 µg of aRNA for each sample was used for sequencing using 2x100 bp non-stranded reads on the Illumina HiSeq 2000 platform, as detailed in Sundell et al. (2015). Raw RNA-Seq data was uploaded to the European Nucleotide Archive (ENA, http://www.ebi.ac.uk/ena/): accession number ERP016242.

RNA-Seq preprocessing

The RNA-Seq data was analyzed using a previously developed pipeline for quality control, read mapping and expression quantification (Delhomme et al., 2014). Briefly, we used FastQC for quality control (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), STAR for read mapping (Dobin et al., 2013), HTSeq for read counts (Anders et al., 2014) and

DESeq for (variance stabilized) gene expression values (Love et al., 2014). The *P. trichocarpa* v3.0 genome and protein-coding gene annotations were used (phytozome.org, v10) (Goodstein et al., 2012). Samples T4-05, T4-09, T4-20 and T4-28 showed low reads counts (Table S12A), while T3-17 was a clear outlier in a principal component analysis (PCA) of all samples (Table S12B). T4-28 was the last sample of T4 and was removed in subsequent analysis. The other four samples were replaced by the average expression values of the two flanking samples. PCA was performed using the R function *prcomp*.

To investigate whether the four trees were clonal replicates or not, a genotype test was performed using Single Nucleotide Polymorphisms (SNPs) in all genes from chromosome 1 in the *P. tricocarpa* genome assembly. RNA-Seq reads from the four replicates, as well as for four different genotypes from an independent experiment (control), were merged using samtools-1.3.1 mpileup (-d100000). SNPs were called using bcftools-1.3.1 call (-v -c). A PCA plot was created from the resulting variant call format file (vcf; Li, 2011) using the R package SNPRelate-1.6.4 (Zheng et al., 2012). From the PCA plot it was clear that the four trees were clonal replicates (Figure S6).

We also developed a pipeline for annotating novel protein coding genes and long intergenic non-coding RNAs (lincRNA) using a combination of establish annotation tools. We used PASA (Haas et al. (2003), 2.0.3, default settings) to construct a reference database based on three assemblies: one *de novo* transcript assembly using the Trinity *de novo* pipeline (Grabherr et al. (2011), r20140717, settings: --min_kmer_cov 1 --normalize_reads --normalize_by_read_set) and two genome guided assemblies using Trinity (Grabherr et al. (2011), r20140717, settings: --genome --genome_guided_use_bam --genome_guided_max_intron 11000 --min_kmer_cov 1) and cufflinks (Trapnell et al. (2010), 2.2.1, settings: --l 24 ibrary-type fr-unstranded -I 15000 --no-faux-reads). PASA reported 59 novel protein coding genes as well as 21,938 Expressed Sequence Tags (ESTs) assemblies with a CDS shorter than 40% of the mRNA length. To identify intergenic RNAs, we removed any region overlapping known *P. trichocarpa* genes (using bedtools (Quinlan and Hall, 2010), v2.19.1, settings: subtract -A). This retained 53 protein-coding genes and 13,995 ESTs. Potential frame-shift errors were identified using frameDP (Gouzy et al. (2009), 1.2.2, default settings) and 672 full length sequences, with CDSs longer than 40% of the mRNA length after correction, changed status to novel protein-coding genes. 816 ESTs with only a start coding, only a stop codon or neither a start nor a stop codon were classified as fragments. The remaining ESTs (12,507 assemblies) were classified as lincRNAs.

Genes were classified as expressed if the variance stabilized gene expression value was above 3.0 in at least two samples in at least three of the four replicate trees. Furthermore, we only considered novel genes longer than 200 bp (see e.g. Ulitsky (2016)). This resulted in 28,294 annotated protein-coding genes, 78 novel protein-coding genes, 567 lincRNAs and 307 fragments with expression during aspen wood development.

Expression analysis

Samples were clustered using Euclidean distance in the R (R-Core-Team, 2012) function *hclust*. Genes were scaled and clustered using Pearson correlation. Ward's method was used in both cases. Dendrograms and heatmaps were generated using the R function *heatmap.2* in the *gplots* library. Samples were reordered within the dendrogram to best match the order in which they were sampled using the R function *reorder*. Functional enrichment was tested using Fishers exact test and false discovery rate corrected. Annotations were downloaded from Phytozome (Goodstein et al., 2012).

To group paralogous genes, the primary protein sequence of 41,335 *P. trichocarpa* genes (Phytozome, Goodstein et al. (2012)) were compared to each other by an all-against-all BLASTP following by a Markov Clustering algorithm (TribeMCL). Genomic homology was detected using i-ADHoRe 3.0 (Proost et al., 2012), using the following settings: alignment method gg4, gap size 30, cluster gap 35, tandem gap 10, q value 0.85, prob cutoff 0.001, anchor points 5, multiple hypothesis correction FDR and level 2 only true. Gene cluster pairs with significantly many shared paralog-pairs (i.e. with one gene in each cluster) were found by comparing the clustering to 10,000 randomly generating clusterings. A circus plot was created using the R function *circos* in the *circlize* library.

Co-expression networks were inferred using mutual information (MI) and context likelihood of relatedness (CLR, Faith et al. (2007)). The CLR algorithm computes a Z-score for each gene pair by using the MI value to all other genes as a null model. A co-expression network was constructed by linking all genes, including annotated protein-coding genes, novel protein-coding genes and lncRNAs, with a Z-score above a given threshold. Our network approached scale freeness (i.e. approached the degree distribution of a random scale-free network generated by the Barabási–Albert model (Barabasi and Albert, 1999)) at a Z-score threshold of five, with higher thresholds not resulting in better approximations, and we therefore used this threshold for most of the analysis in this paper. Links based on MI values below 0.25 were deemed spurious and removed. We computed gene centrality for each gene in the network, including degree (number of neighbors), average neighborhood degree (the average degree of the neighbors), betweenness (the probability that the gene is part of the shortest network path between two arbitrary genes) and closeness (the shortest distance between a gene and all other genes in the network). The expression specificity score of a gene was calculated as the highest observed ratio between the average expression within and outside a zone of consecutive samples. All zones containing from three to 10 samples were considered and the final score was calculated as the average of the highest score from each replicate tree. The clustered network was generated as explained in the figure text of Figure 3.

AspWood

A web resource was built to allow the community easy access to the expression data (AspWood, http://aspwood.popgenie.org). The user can start with an existing list of genes, create a new list by text search or select one of the pre-

computed clusters. The tool will generate a page for the selected gene list with modules showing the expression profiles, the co-expression network, gene information, functional enrichments and the heatmap. The co-expression network is interactive and allows the user to explore the data further. The gene list can be exported and used in tools throughout the PlantGenIE platform (http://plantgenie.org, Sundell et al. (2015)). AspWood is built with HTML5 and JavaScript for user interfaces, and PhP and Python for more advanced server side functions.

Declarations

Acknowledgments

Thanks to Kjell Olofsson for performing the cryosectioning and to Nicolas Delhomme for advice on RNA-Seq preprocessing.

Funding

This study was funded by the Swedish Foundation for Strategic Research (SFF) and the Swedish Governmental Agency for Innovation Systems (VINNOVA) through the UPSC Berzelii Centre for Forest Biotechnology.

Availability of data and materials

Raw RNA-Seq reads are available at the European Nucleotide Archive (ENA, http://www.ebi.ac.uk/ena/): accession number ERP016242. Processed data is available through the interactive AspWood web resource: http://aspwood.popgenie.org.

Authors' contributions

NRS, BS and TRH conceived and designed the experiment. MK prepared RNA for sequencing. DS implemented AspWood (with help from CM), and preformed the RNA-Seq preprocessing, the novel gene analysis and co-expression network and enrichment analysis. TRH performed the cluster analysis. NRS, EJM, MK, CJ, VK, ON, HT, EP, UF, TN, BS and TRH analyzed and interpreted the data. All authors contributed text to the manuscript, and read and approved the final version.

Competing interests

The authors declare that they have no competing interest.

Additional figures and tables

Additional figures

Figure S1.
Plots showing the expression diversity across the sampled series. Sample colors indicate the sample clusters from Figure 1B. (A) Box plots showing the expression distribution in each sample. The boxes indicate upper/lower quartiles (i.e. 25% of the genes are expressed higher/lower than the box) with the horizontal lines marking the median (i.e. 50% of the genes are expressed higher/lower than this value). The lines extending vertically from the boxes (whiskers) indicate the maximum/minimum values excluding outliers. Outliers (i.e. genes expressed more/less than 1.5 times the upper/lower quartile) are plotted as open circles. (B, C) The number of genes in each sample expressed above a threshold of 3 and 15, respectively. (D) The number of genes in each sample accounting for 90% of the expression in that sample.

Figure S2.
The sample clustering in Figure 1B with visible sample names (see Table S2).

Figure S3.
Marker genes. Expression is shown with (A) the variance stabilized transformation (VST) and (B) scaled counts per million (CPM, calculated as 2^{VST}, scaled: mean centered and normalized by the standard deviation of each gene). SUS6/Potri.004G081300: *A. thaliana* sucrose synthase (SUS) genes SUS5 and SUS6 are known to be phloem localized (Barratt et al., 2009). The *Populus* orthologs of SUS5 and SUS6 peaked in the two or three outermost sections of the single section series covering the cambial meristem, thus marking a phloem identity to these samples. The pooled phloem sample, consisting of six to seven section showed lower levels indicating that SUS5 and SUS6 expression decreased in the older phloem tissues. Expression was shown for SUS6. CDC2.2/Potri.016G142800: The dividing meristem was marked by a set of cyclins typically known to be involved in the different stages of cell cycling. These markers indicated that sample cluster ***i*** and ***ii*** from the hierarchical clustering was split in the middle of the meristem. Expression was shown for CDC2.2. *PtEXPA1*/Potri.001G240900: Cell expansion was marked by the *Populus* alpha expansin (Gray-Mitsumune et al., 2004). The *PtEXPA1* expression marked the broader region covering both the phloem and xylem side, and decreased sharply at the border to sample cluster ***iii***. *PtCesA8-B*/Potri.004G059600: *CesA4, 7* and *8* are well-described hallmark genes for secondary wall formation in xylem cells (Kumar et al., 2009). The corresponding transcripts in aspen showed two distinct peaks, one minor on the phloem side marking secondary wall formation in phloem, and another major peak on the xylem side marking the bulk of secondary wall formation in xylem vessels, fibers and rays. Expression was shown for *PtCesA8-B*. BFN1.1/Potri.011G044500: Vessel cells undergo cell death before fiber cells. Xylem specific proteases and nucleases are responsible for the autolysis of xylem cells following cell death were selected to mark the timing of the cell death of fibers and vessel elements (reviewed by Escamez and Tuominen (2014)).

In aspen, an ortholog of the *A. thaliana* bifunctional nuclease 1 (BFN1) protein showed two peaks of expression; one appearing at the end of sample cluster ***iii*** and the other at the end of sample cluster ***iv***. This corresponds well with the locations of cell death, as estimated on the basis of the viability assays of the xylem tissues (TableS2). Thus, the sharp decrease in the expression of *BFN1* marks the border of living vessel elements and fibers, respectively.

Figure S4.
The hierarchical clustering from Figure 1BC, but with samples ordered according to sampling order for each of the four replicate trees and not according to expression similarity. The four sample clusters identified in Figure 1B are indicated by the color bar above the heatmap.

Figure S5.
Comparison of expression domains of proposed direct targets of *PtSND1* paralogs in wood-forming tissues. The heatmap shows expression profiles of PtSND1s (in green) as well as proposed direct targets of these identified in previous studies (in blue (Ye and Zhong, 2015) and black (Lin et al., 2013)). Genes proposed as targets by both previous studies are in red.

Figure S6.
Genotyping. RNA-Seq reads from the four AspWood replicates, as well as from four different genotypes from an independent experiment (controls), were used to call SNPs for all genes on Chromosome 1 of the *P. trichocarpa* genome. A Principle Component Analysis (PCA) showed that the four AspWood trees (turquoise) were clonal replicates and that the controls (red) were not.

Additional tables

Table S1. Information about the sampled trees including height, diameter, age, and sampling height.

Table S2. Anatomical characterization of each section in the sample series, and description of pooling for RNA-Seq analysis. The colors indicate sample clusters according to Figure 1B. Cell viability measurements are also indicated (purple bars).

Table S3. Expression values for the 28,294 expressed, annotated protein-coding genes (version 3 of the *P. trichocarpa* genome).

Table S4. Expression values for the 78 expressed, novel protein-coding genes identified in this study. Chromosomal positions in the *P. trichocarpa* genome is also indicated.

Table S5. Expression values for the 567 expressed, long non-coding RNAs identified in this study. Chromosomal positions in the *P. trichocarpa* genome is also indicated.

Table S6. Expression values for the 307 expressed, transcript fragments identified in this study. Chromosomal positions in the *P. trichocarpa* genome is also indicated.

Table S7. All 28,294 expressed, annotated protein-coding genes with cluster assignments according to the hierarchical clustering. Functional annotations are from www.phytozome.net.

Table S8. Network centrality statistics for the genes in the co-expression network at a co-expression threshold of five: 13,882 annotated genes (Sheet named *Annotated*), 25 novel coding genes (*Novel coding*), 205 novel long non-coding RNAs (*Novel non-coding*) and 87 transcript fragments (*Novel fragments*). Each sheet includes columns for network statistics, expression specificity (see Methods), Pearson correlations to the mean expression profiles of the gene clusters obtained by hierarchical clustering and annotations (www.phytozome.net). A separate sheet contains the same information for genes selected as representatives in the clustered network (*Annotated representative*). Finally, there is a sheet providing statistical associations observed in the network (*Statistics*) and a sheet providing statistics on BLAST hit with E-value < 1E-10 to other species (*BLAST*).

Table S9. Central genes and enriched gene functions in the 41 network clusters.

Table S10. Complete lists of genes for the different biological stories.

Table S11. Analysis of WGD paralogs. (A) Expression correlation for all paralog pairs, where at least one member of the pair was expressed in the section series. (B) Numbers of pairs (with p-values) shared between the clusters. Table with the number of diverged pairs after applying different correlation thresholds.

Table S12. RNA-seq data quality. (A) Read counts for each sample including the number of raw reads and the number of reads after filtering, trimming and alignment (in millions). (B) Principle Component Analysis (PCA) of all samples based on the variance stabilized expression data.

References

Anders, S., Pyl, P.T., and Huber, W. (2014). HTSeq; A Python framework to work with high-throughput sequencing data.

Aspeborg, H., Schrader, J., Coutinho, P.M., Stam, M., Kallas, A., Djerbi, S., Nilsson, P., Denman, S., Amini, B., Sterky, F., Master, E., Sandberg, G., Mellerowicz, E., Sundberg, B., Henrissat, B., and Teeri, T.T. (2005). Carbohydrate-active enzymes involved in the secondary cell wall biogenesis in hybrid aspen. Plant Physiol **137,** 983-997.

Atmodjo, M.A., Sakuragi, Y., Zhu, X., Burrell, A.J., Mohanty, S.S., Atwood, J.A., 3rd, Orlando, R., Scheller, H.V., and Mohnen, D. (2011). Galacturonosyltransferase (GAUT)1 and GAUT7 are the core of a plant cell wall pectin biosynthetic homogalacturonan:galacturonosyltransferase complex. Proc Natl Acad Sci U S A **108,** 20225-20230.

Barabasi, A.L., and Albert, R. (1999). Emergence of scaling in random networks. Science **286,** 509-512.

Barnett, J.R. (1981). Secondary xylem cell development. In Xylem Cell Development, J.R. Barnett, ed (London, England: Castle House), pp. 47-95.

Barratt, D.H.P., Derbyshire, P., Findlay, K., Pike, M., Wellner, N., Lunn, J., Feil, R., Simpson, C., Maule, A.J., and Smith, A.M. (2009). Normal growth of Arabidopsis requires cytosolic invertase but not sucrose synthase. P Natl Acad Sci USA **106,** 13124-13129.

Barros, J., Serk, H., Granlund, I., and Pesquet, E. (2015). The cell biology of lignification in higher plants. Ann Bot **115,** 1053-1074.

Biswal, A.K., Soeno, K., Gandla, M.L., Immerzeel, P., Pattathil, S., Lucenius, J., Serimaa, R., Hahn, M.G., Moritz, T., Jonsson, L.J., Israelsson-Nordstrom, M., and Mellerowicz, E.J. (2014). Aspen pectate lyase PtxtPL1-27 mobilizes matrix polysaccharides from woody tissues and improves saccharification yield. Biotechnol Biofuels **7,** 11.

Boerjan, W., Ralph, J., and Baucher, M. (2003). Lignin biosynthesis. Annu Rev Plant Biol **54,** 519-546.

Bollhoner, B., Prestele, J., and Tuominen, H. (2012). Xylem cell death: emerging understanding of regulation and function. J Exp Bot **63,** 1081-1094.

Bollhoner, B., Zhang, B., Stael, S., Denance, N., Overmyer, K., Goffner, D., Van Breusegem, F., and Tuominen, H. (2013). Post mortem function of AtMC9 in xylem vessel elements. New Phytol **200,** 498-510.

Bonke, M., Thitamadee, S., Mahonen, A.P., Hauser, M.T., and Helariutta, Y. (2003). APL regulates vascular tissue identity in Arabidopsis. Nature **426,** 181-186.

Bourquin, V., Nishikubo, N., Abe, H., Brumer, H., Denman, S., Eklund, M., Christiernin, M., Teeri, T.T., Sundberg, B., and Mellerowicz, E.J. (2002). Xyloglucan endotransglycosylases have a function during the formation of secondary cell walls of vascular tissues. Plant Cell **14,** 3073-3088.

Carroll, A., Mansoori, N., Li, S., Lei, L., Vernhettes, S., Visser, R.G., Somerville, C., Gu, Y., and Trindade, L.M. (2012). Complexes with mixed primary and secondary cellulose synthases are functional in Arabidopsis plants. Plant Physiol **160,** 726-737.

Cavalier, D.M., and Keegstra, K. (2006). Two xyloglucan xylosyltransferases catalyze the addition of multiple xylosyl residues to cellohexaose. J Biol Chem **281,** 34197-34207.

Chou, Y.H., Pogorelko, G., and Zabotina, O.A. (2012). Xyloglucan xylosyltransferases XXT1, XXT2, and XXT5 and the glucan synthase CSLC4 form Golgi-localized multiprotein complexes. Plant Physiol **159,** 1355-1366.

Christensen, J.H., Bauw, G., Welinder, K.G., Van Montagu, M., and Boerjan, W. (1998). Purification and characterization of peroxidases correlated with lignification in poplar xylem. Plant Physiol **118,** 125-135.

Courtois-Moreau, C.L., Pesquet, E., Sjodin, A., Muniz, L., Bollhoner, B., Kaneda, M., Samuels, L., Jansson, S., and Tuominen, H. (2009). A unique program for cell death in xylem fibers of Populus stem. Plant J **58,** 260-274.

Dai, X., Hu, Q., Cai, Q., Feng, K., Ye, N., Tuskan, G.A., Milne, R., Chen, Y., Wan, Z., Wang, Z., Luo, W., Wang, K., Wan, D., Wang, M., Wang, J., Liu, J., and Yin, T. (2014). The willow genome and divergent evolution from poplar after the common genome duplication. Cell Res **24,** 1274-1277.

De Rybel, B., Moller, B., Yoshida, S., Grabowicz, I., Barbier de Reuille, P., Boeren, S., Smith, R.S., Borst, J.W., and Weijers, D. (2013). A bHLH complex controls embryonic vascular tissue establishment and indeterminate growth in Arabidopsis. Developmental cell **24,** 426-437.

De Rybel, B., Adibi, M., Breda, A.S., Wendrich, J.R., Smit, M.E., Novak, O., Yamaguchi, N., Yoshida, S., Van Isterdael, G., Palovaara, J., Nijsse, B., Boekschoten, M.V., Hooiveld, G., Beeckman, T., Wagner, D., Ljung, K., Fleck, C., and Weijers, D. (2014). Plant development. Integration of growth and patterning during vascular tissue formation in Arabidopsis. Science **345,** 1255215.

Delhomme, N., Mähler, N., Schiffthaler, B., Sundell, D., Mannapperuma, C., Hvidsten, T., and Street, N. (2014). Guidelines for RNA-Seq data analysis. EpiGeneSys Protocol.

Depuydt, S., Rodriguez-Villalon, A., Santuari, L., Wyser-Rmili, C., Ragni, L., and Hardtke, C.S. (2013). Suppression of Arabidopsis protophloem differentiation and root meristem growth by CLE45 requires the receptor-like kinase BAM3. Proc Natl Acad Sci U S A **110,** 7074-7079.

Derbyshire, P., Menard, D., Green, P., Saalbach, G., Buschmann, H., Lloyd, C.W., and Pesquet, E. (2015). Proteomic Analysis of Microtubule Interacting Proteins over the Course of Xylem Tracheary Element Formation in Arabidopsis. Plant Cell **27,** 2709-2726.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics **29,** 15-21.

Egelund, J., Petersen, B.L., Motawia, M.S., Damager, I., Faik, A., Olsen, C.E., Ishii, T., Clausen, H., Ulvskov, P., and Geshi, N. (2006). Arabidopsis thaliana RGXT1 and RGXT2 encode Golgi-localized (1,3)-alpha-D-xylosyltransferases involved in the synthesis of pectic rhamnogalacturonan-II. Plant Cell **18,** 2593-2607.

Emery, J.F., Floyd, S.K., Alvarez, J., Eshed, Y., Hawker, N.P., Izhaki, A., Baum, S.F., and Bowman, J.L. (2003). Radial patterning of Arabidopsis shoots by class III HD-ZIP and KANADI genes. Current biology : CB **13,** 1768-1774.

Escamez, S., and Tuominen, H. (2014). Programmes of cell death and autolysis in tracheary elements: when a suicidal cell arranges its own corpse removal. J Exp Bot **65,** 1313-1321.

Espinosa-Ruiz, A., Saxena, S., Schmidt, J., Mellerowicz, E., Miskolczi, P., Bako, L., and Bhalerao, R.P. (2004). Differential stage-specific regulation of cyclin-dependent kinases during cambial dormancy in hybrid aspen. Plant Journal **38,** 603-615.

Etchells, J.P., and Turner, S.R. (2010). The PXY-CLE41 receptor ligand pair defines a multifunctional pathway that controls the rate and orientation of vascular cell division. Development **137,** 767-774.

Evert, R.F., and Eichhorn, S.E. (2006). In Esau's Plant Anatomy (Willey & Sons, Inc), pp. 407-425.

Faith, J.J., Hayete, B., Thaden, J.T., Mogno, I., Wierzbowski, J., Cottarel, G., Kasif, S., Collins, J.J., and Gardner, T.S. (2007). Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. PLoS biology **5,** e8.

Fujii, T., Harada, H., and Saiki, H. (1981). Ultrastructure of 'amorphous layer' in xylem parenchyma cell wall of angiosperm species. Mokuzai Gakkaishi **27,** 149–156.

Fukuda, H. (2016). Signaling, transcriptional regulation, and asynchronous pattern formation governing plant xylem development. Proc Jpn Acad Ser B Phys Biol Sci **92,** 98-107.

Furuta, K.M., Yadav, S.R., Lehesranta, S., Belevich, I., Miyashima, S., Heo, J.O., Vaten, A., Lindgren, O., De Rybel, B., Van Isterdael, G., Somervuo, P., Lichtenberger, R., Rocha, R., Thitamadee, S., Tahtiharju, S., Auvinen, P., Beeckman, T., Jokitalo, E., and Helariutta, Y. (2014). Plant development. Arabidopsis NAC45/86 direct sieve element morphogenesis culminating in enucleation. Science **345,** 933-937.

Gerber, L., Zhang, B., Roach, M., Rende, U., Gorzsas, A., Kumar, M., Burgert, I., Niittyla, T., and Sundberg, B. (2014). Deficient sucrose synthase activity in developing wood does not specifically affect cellulose biosynthesis, but causes an overall decrease in cell wall polymers. New Phytol **203,** 1220-1230.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., and Rokhsar, D.S. (2012). Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res **40,** D1178-1186.

Goodwin, S., McPherson, J.D., and McCombie, W.R. (2016). Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet **17,** 333-351.

Gouzy, J., Carrere, S., and Schiex, T. (2009). FrameDP: sensitive peptide detection on noisy matured sequences. Bioinformatics **25,** 670-671.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., and Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol **29,** 644-652.

Gray-Mitsumune, M., Blomquist, K., McQueen-Mason, S., Teeri, T.T., Sundberg, B., and Mellerowicz, E.J. (2008). Ectopic expression of a wood-abundant expansin PttEXPA1 promotes cell expansion in primary and secondary tissues in aspen. Plant Biotechnol J **6,** 62-72.

Gray-Mitsumune, M., Mellerowicz, E.J., Abe, H., Schrader, J., Winzell, A., Sterky, F., Blomqvist, K., McQueen-Mason, S., Teeri, T.T., and Sundberg, B. (2004). Expansins abundant in secondary xylem belong to subgroup A of the alpha-expansin gene family. Plant Physiol **135,** 1552-1564.

Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith, R.K., Jr., Hannick, L.I., Maiti, R., Ronning, C.M., Rusch, D.B., Town, C.D., Salzberg, S.L., and White, O. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res **31,** 5654-5666.

Harholt, J., Jensen, J.K., Sorensen, S.O., Orfila, C., Pauly, M., and Scheller, H.V. (2006). ARABINAN DEFICIENT 1 is a putative arabinosyltransferase involved in biosynthesis of pectic arabinan in Arabidopsis. Plant Physiol **140,** 49-58.

Herbette, S., Bouchet, B., Brunel, N., Bonnin, E., Cochard, H., and Guillon, F. (2015). Immunolabelling of intervessel pits for polysaccharides and lignin helps in understanding their hydraulic properties in Populus tremula x alba. Ann Bot **115,** 187-199.

Hermansen, R.A., Hvidsten, T.R., Sandve, S.R., and Liberles, D.A. (2016). Extracting functional trends from whole genome duplication events using comparative genomics. Biol Proced Online **18,** 11.

Herrero, J., Fernandez-Perez, F., Yebra, T., Novo-Uzal, E., Pomar, F., Pedreno, M.A., Cuello, J., Guera, A., Esteban-Carrasco, A., and Zapata, J.M. (2013). Bioinformatic and functional characterization of the basic peroxidase 72 from Arabidopsis thaliana involved in lignin biosynthesis. Planta **237,** 1599-1612.

Hertzberg, M., Aspeborg, H., Schrader, J., Andersson, A., Erlandsson, R., Blomqvist, K., Bhalerao, R., Uhlen, M., Teeri, T.T., Lundeberg, J., Sundberg, B., Nilsson, P., and Sandberg, G. (2001). A transcriptional roadmap to wood formation. Proc Natl Acad Sci U S A **98,** 14732-14737.

Hill, J.L., Jr., Hammudi, M.B., and Tien, M. (2014). The Arabidopsis cellulose synthase complex: a proposed hexamer of CESA trimers in an equimolar stoichiometry. Plant Cell **26,** 4834-4842.

Hirakawa, Y., Shinohara, H., Kondo, Y., Inoue, A., Nakanomyo, I., Ogawa, M., Sawa, S., Ohashi-Ito, K., Matsubayashi, Y., and Fukuda, H. (2008). Non-cell-autonomous control of vascular stem cell fate by a CLE peptide/receptor system. Proc Natl Acad Sci U S A **105,** 15208-15213.

Ilegems, M., Douet, V., Meylan-Bettex, M., Uyttewaal, M., Brand, L., Bowman, J.L., and Stieger, P.A. (2010). Interplay of auxin, KANADI and Class III HD-ZIP transcription factors in vascular tissue formation. Development **137,** 975-984.

Immanen, J., Nieminen, K., Smolander, O.P., Kojima, M., Alonso Serra, J., Koskinen, P., Zhang, J., Elo, A., Mahonen, A.P., Street, N., Bhalerao, R.P., Paulin, L., Auvinen, P., Sakakibara, H., and Helariutta, Y. (2016).

Cytokinin and Auxin Display Distinct but Interconnected Distribution and Signaling Profiles to Stimulate Cambial Activity. Curr Biol **26,** 1990-1997.

Ito, J., and Fukuda, H. (2002). ZEN1 is a key enzyme in the degradation of nuclear DNA during programmed cell death of tracheary elements. Plant Cell **14,** 3201-3211.

Jensen, J.K., Johnson, N.R., and Wilkerson, C.G. (2014). Arabidopsis thaliana IRX10 and two related proteins from psyllium and Physcomitrella patens are xylan xylosyltransferases. Plant J **80,** 207-215.

Johnsson, C., and Fischer, U. (2016). Cambial stem cells and their niche. Plant Sci **252,** 239-245.

Kubo, M., Udagawa, M., Nishikubo, N., Horiguchi, G., Yamaguchi, M., Ito, J., Mimura, T., Fukuda, H., and Demura, T. (2005). Transcription switches for protoxylem and metaxylem vessel formation. Genes Dev **19,** 1855-1860.

Kumar, M., Thammannagowda, S., Bulone, V., Chiang, V., Han, K.H., Joshi, C.P., Mansfield, S.D., Mellerowicz, E., Sundberg, B., Teeri, T., and Ellis, B.E. (2009). An update on the nomenclature for the cellulose synthase genes in Populus. Trends Plant Sci **14,** 248-254.

Larson, P.R. (1994). The Vascular Cambium: Development and Structure. (Berlin: Springer-Verlag).

Lee, C., Teng, Q., Zhong, R., and Ye, Z.H. (2011). Molecular dissection of xylan biosynthesis during wood formation in poplar. Mol Plant **4,** 730-747.

Li, Y., Kajita, S., Kawai, S., Katayama, Y., and Morohoshi, N. (2003). Down-regulation of an anionic peroxidase in transgenic aspen and its effect on lignin characteristics. J Plant Res **116,** 175-182.

Lin, Y.C., Li, W., Sun, Y.H., Kumari, S., Wei, H., Li, Q., Tunlaya-Anukit, S., Sederoff, R.R., and Chiang, V.L. (2013). SND1 transcription factor-directed quantitative functional hierarchical genetic regulatory network in wood formation in Populus trichocarpa. Plant Cell **25,** 4324-4341.

Liwanag, A.J., Ebert, B., Verhertbruggen, Y., Rennie, E.A., Rautengarten, C., Oikawa, A., Andersen, M.C., Clausen, M.H., and Scheller, H.V. (2012). Pectin biosynthesis: GALS1 in Arabidopsis thaliana is a beta-1,4-galactan beta-1,4-galactosyltransferase. Plant Cell **24,** 5024-5036.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol **15,** 550.

Lu, S., Li, Q., Wei, H., Chang, M.J., Tunlaya-Anukit, S., Kim, H., Liu, J., Song, J., Sun, Y.H., Yuan, L., Yeh, T.F., Peszlen, I., Ralph, J., Sederoff, R.R., and Chiang, V.L. (2013). Ptr-miR397a is a negative regulator of laccase genes affecting lignin content in Populus trichocarpa. Proc Natl Acad Sci U S A **110,** 10848-10853.

McCarthy, R.L., Zhong, R., Fowler, S., Lyskowski, D., Piyasena, H., Carleton, K., Spicer, C., and Ye, Z.H. (2010). The poplar MYB transcription factors, PtrMYB3 and PtrMYB20, are involved in the regulation of secondary wall biosynthesis. Plant Cell Physiol **51,** 1084-1090.

Mellerowicz, E.J., and Sundberg, B. (2008). Wood cell walls: biosynthesis, developmental dynamics and their implications for wood properties. Curr Opin Plant Biol **11,** 293-300.

Mellerowicz, E.J., and Gorshkova, T.A. (2012). Tensional stress generation in gelatinous fibres: a review and possible mechanism based on cell-wall structure and composition. J Exp Bot **63,** 551-565.

Mellerowicz, E.J., Baucher, M., Sundberg, B., and Boerjan, W. (2001). Unravelling cell wall formation in the woody dicot stem. Plant Mol Biol **47,** 239-274.

Moreau, C., Aksenov, N., Lorenzo, M.G., Segerman, B., Funk, C., Nilsson, P., Jansson, S., and Tuominen, H. (2005). A genomic approach to investigate developmental cell death in woody tissues of Populus trees. Genome Biol **6,** R34.

Mortimer, J.C., Faria-Blanc, N., Yu, X., Tryfona, T., Sorieul, M., Ng, Y.Z., Zhang, Z., Stott, K., Anders, N., and Dupree, P. (2015). An unusual xylan in Arabidopsis primary cell walls is synthesised by GUX3, IRX9L, IRX10L and IRX14. Plant J **83,** 413-426.

Murakami, R., Funada, R., Sano, Y., and Ohtani, J. (1999). The Differentiation of Contact Cells and Isolation Cells in the Xylem Ray Parenchyma of Populus maximowiczii. Annals of Botany **84,** 429-435.

Mutwil, M., Klie, S., Tohge, T., Giorgi, F.M., Wilkins, O., Campbell, M.M., Fernie, A.R., Usadel, B., Nikoloski, Z., and Persson, S. (2011). PlaNet: combined sequence and expression comparisons across plant networks derived from seven species. Plant Cell **23,** 895-910.

Nakaba, S., Begum, S., Yamagishi, Y., Jin, H.-O., Kubo, T., and Funada, R. (2012). Differences in the timing of cell death, differentiation and function among three different types of ray parenchyma cells in the hardwood Populus sieboldii × P. grandidentata. Trees **26,** 743.

Netotea, S., Sundell, D., Street, N.R., and Hvidsten, T.R. (2014). ComPlEx: conservation and divergence of co-expression networks in A. thaliana, Populus and O. sativa. BMC Genomics **15,** 106.

Ohashi-Ito, K., Saegusa, M., Iwamoto, K., Oda, Y., Katayama, H., Kojima, M., Sakakibara, H., and Fukuda, H. (2014). A bHLH complex activates vascular cell division via cytokinin action in root apical meristem. Current biology : CB **24,** 2053-2058.

Oikawa, A., Lund, C.H., Sakuragi, Y., and Scheller, H.V. (2013). Golgi-localized enzyme complexes for plant cell wall biosynthesis. Trends Plant Sci **18,** 49-58.

Pesquet, E., Zhang, B., Gorzsas, A., Puhakainen, T., Serk, H., Escamez, S., Barbier, O., Gerber, L., Courtois-Moreau, C., Alatalo, E., Paulin, L., Kangasjarvi, J., Sundberg, B., Goffner, D., and Tuominen, H. (2013). Non-cell-autonomous postmortem lignification of tracheary elements in Zinnia elegans. Plant Cell **25,** 1314-1328.

Proost, S., Fostier, J., De Witte, D., Dhoedt, B., Demeester, P., Van de Peer, Y., and Vandepoele, K. (2012). i-ADHoRe 3.0--fast and sensitive detection of genomic homology in extremely large data sets. Nucleic Acids Res **40,** e11.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics **26,** 841-842.

R-Core-Team. (2012). R: A Language and Environment for Statistical Computing (Vienna, 780 Austria OR - R Foundation for Statistical Computing).

Ranocha, P., McDougall, G., Hawkins, S., Sterjiades, R., Borderies, G., Stewart, D., Cabanes-Macheteau, M., Boudet, A.M., and Goffner, D. (1999).

Biochemical characterization, molecular cloning and expression of laccases - a divergent gene family - in poplar. Eur J Biochem **259,** 485-495.

Ratke, C., Pawar, P.M., Balasubramanian, V.K., Naumann, M., Duncranz, M.L., Derba-Maceluch, M., Gorzsas, A., Endo, S., Ezcurra, I., and Mellerowicz, E.J. (2015). Populus GT43 family members group into distinct sets required for primary and secondary wall xylan biosynthesis and include useful promoters for wood modification. Plant Biotechnol J **13,** 26-37.

Rodgers-Melnick, E., Mane, S.P., Dharmawardhana, P., Slavov, G.T., Crasta, O.R., Strauss, S.H., Brunner, A.M., and Difazio, S.P. (2012). Contrasting patterns of evolution following whole genome versus tandem duplication events in Populus. Genome Res **22,** 95-105.

Rodriguez-Villalon, A., Gujas, B., Kang, Y.H., Breda, A.S., Cattaneo, P., Depuydt, S., and Hardtke, C.S. (2014). Molecular genetic framework for protophloem formation. Proc Natl Acad Sci U S A **111,** 11551-11556.

Rosin, F.M., and Kramer, E.M. (2009). Old dogs, new tricks: regulatory evolution in conserved genetic modules leads to novel morphologies in plants. Dev Biol **332,** 25-35.

Ruzicka, K., Ursache, R., Hejatko, J., and Helariutta, Y. (2015). Xylem development - from the cradle to the grave. New Phytol **207,** 519-535.

Sasaki, S., Nishida, T., Tsutsumi, Y., and Kondo, R. (2004). Lignin dehydrogenative polymerization mechanism: a poplar cell wall peroxidase directly oxidizes polymer lignin and produces in vitro dehydrogenative polymer rich in beta-O-4 linkage. FEBS Lett **562,** 197-201.

Sasaki, S., Nonaka, D., Wariishi, H., Tsutsumi, Y., and Kondo, R. (2008). Role of Tyr residues on the protein surface of cationic cell-wall-peroxidase (CWPO-C) from poplar: potential oxidation sites for oxidative polymerization of lignin. Phytochemistry **69,** 348-355.

Scacchi, E., Salinas, P., Gujas, B., Santuari, L., Krogan, N., Ragni, L., Berleth, T., and Hardtke, C.S. (2010). Spatio-temporal sequence of cross-regulatory events in root meristem growth. Proc Natl Acad Sci U S A **107,** 22734-22739.

Schlereth, A., Moller, B., Liu, W., Kientz, M., Flipse, J., Rademacher, E.H., Schmid, M., Jurgens, G., and Weijers, D. (2010). MONOPTEROS controls embryonic root initiation by regulating a mobile transcription factor. Nature **464,** 913-916.

Schrader, J., Nilsson, J., Mellerowicz, E., Berglund, A., Nilsson, P., Hertzberg, M., and Sandberg, G. (2004). A high-resolution transcript profile across the wood-forming meristem of poplar identifies potential regulators of cambial stem cell identity. Plant Cell **16,** 2278-2292.

Shi, R., Sun, Y.H., Li, Q., Heber, S., Sederoff, R., and Chiang, V.L. (2010). Towards a systems approach for lignin biosynthesis in Populus trichocarpa: transcript abundance and specificity of the monolignol biosynthetic genes. Plant Cell Physiol **51,** 144-163.

Shigeto, J., Itoh, Y., Hirao, S., Ohira, K., Fujita, K., and Tsutsumi, Y. (2015). Simultaneously disrupting AtPrx2, AtPrx25 and AtPrx71 alters lignin content and structure in Arabidopsis stem. J Integr Plant Biol **57,** 349-356.

Song, D., Shen, J., and Li, L. (2010). Characterization of cellulose synthase complexes in Populus xylem differentiation. New Phytol **187,** 777-790.

Sterck, L., Rombauts, S., Jansson, S., Sterky, F., Rouze, P., and Van de Peer, Y. (2005). EST data suggest that poplar is an ancient polyploid. New Phytol **167,** 165-170.

Sundell, D., Mannapperuma, C., Netotea, S., Delhomme, N., Lin, Y.C., Sjodin, A., Van de Peer, Y., Jansson, S., Hvidsten, T.R., and Street, N.R. (2015). The Plant Genome Integrative Explorer Resource: PlantGenIE.org. New Phytol **208,** 1149-1156.

Takata, N., and Taniguchi, T. (2015). Expression divergence of cellulose synthase (CesA) genes after a recent whole genome duplication event in Populus. Planta **241,** 29-42.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol **28,** 511-515.

Truernit, E., Bauby, H., Belcram, K., Barthelemy, J., and Palauqui, J.C. (2012). OCTOPUS, a polarly localised membrane-associated protein, regulates phloem differentiation entry in Arabidopsis thaliana. Development (Cambridge, England) **139,** 1306-1315.

Tuskan, G.A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R.R., Bhalerao, R.P., Blaudez, D., Boerjan, W., Brun, A., Brunner, A., Busov, V., Campbell, M., Carlson, J., Chalot, M., Chapman, J., Chen, G.L., Cooper, D., Coutinho, P.M., Couturier, J., Covert, S., Cronk, Q., Cunningham, R., Davis, J., Degroeve, S., Dejardin, A., Depamphilis, C., Detter, J., Dirks, B., Dubchak, I., Duplessis, S., Ehlting, J., Ellis, B., Gendler, K., Goodstein, D., Gribskov, M., Grimwood, J., Groover, A., Gunter, L., Hamberger, B., Heinze, B., Helariutta, Y., Henrissat, B., Holligan, D., Holt, R., Huang, W., Islam-Faridi, N., Jones, S., Jones-Rhoades, M., Jorgensen, R., Joshi, C., Kangasjarvi, J., Karlsson, J., Kelleher, C., Kirkpatrick, R., Kirst, M., Kohler, A., Kalluri, U., Larimer, F., Leebens-Mack, J., Leple, J.C., Locascio, P., Lou, Y., Lucas, S., Martin, F., Montanini, B., Napoli, C., Nelson, D.R., Nelson, C., Nieminen, K., Nilsson, O., Pereda, V., Peter, G., Philippe, R., Pilate, G., Poliakov, A., Razumovskaya, J., Richardson, P., Rinaldi, C., Ritland, K., Rouze, P., Ryaboy, D., Schmutz, J., Schrader, J., Segerman, B., Shin, H., Siddiqui, A., Sterky, F., Terry, A., Tsai, C.J., Uberbacher, E., Unneberg, P., Vahala, J., Wall, K., Wessler, S., Yang, G., Yin, T., Douglas, C., Marra, M., Sandberg, G., Van de Peer, Y., and Rokhsar, D. (2006). The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science **313,** 1596-1604.

Uggla, C., and Sundberg, B. (2002). Chapter 13. Sampling of Cambial Region Tissues for High Resolution Analysis. In Wood Formation in Trees (CRC Press), pp. 215–228.

Uggla, C., Moritz, T., Sandberg, G., and Sundberg, B. (1996). Auxin as a positional signal in pattern formation in plants. Proc Natl Acad Sci U S A **93,** 9282-9286.

Ulitsky, I. (2016). Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. Nat Rev Genet **17,** 601-614.

Urbanowicz, B.R., Pena, M.J., Moniz, H.A., Moremen, K.W., and York, W.S. (2014). Two Arabidopsis proteins synthesize acetylated xylan in vitro. Plant J **80,** 197-206.

Wullschleger, S.D., Weston, D.J., DiFazio, S.P., and Tuskan, G.A. (2013). Revisiting the sequencing of the first tree genome: Populus trichocarpa. Tree physiology **33,** 357-364.

Ye, Z.H., and Zhong, R. (2015). Molecular control of wood formation in trees. J Exp Bot **66,** 4119-4131.

Yoo, M.J., Liu, X., Pires, J.C., Soltis, P.S., and Soltis, D.E. (2014). Nonadditive gene expression in polyploids. Annu Rev Genet **48,** 485-517.

Zabotina, O.A., van de Ven, W.T., Freshour, G., Drakakaki, G., Cavalier, D., Mouille, G., Hahn, M.G., Keegstra, K., and Raikhel, N.V. (2008). Arabidopsis XXT5 gene encodes a putative alpha-1,6-xylosyltransferase that is involved in xyloglucan biosynthesis. Plant J **56,** 101-115.

Zeng, W., Lampugnani, E.R., Picard, K.L., Song, L., Wu, A.M., Farion, I.M., Zhao, J., Ford, K., Doblin, M.S., and Bacic, A. (2016). Asparagus IRX9, IRX10, and IRX14A Are Components of an Active Xylan Backbone Synthase Complex that Forms in the Golgi Apparatus. Plant Physiol **171,** 93-109.

Zhang, W., Landback, P., Gschwend, A.R., Shen, B., and Long, M. (2015). New genes drive the evolution of gene interaction networks in the human and mouse genomes. Genome Biol **16,** 202.

Zhao, Q., Nakashima, J., Chen, F., Yin, Y., Fu, C., Yun, J., Shao, H., Wang, X., Wang, Z.Y., and Dixon, R.A. (2013). Laccase is necessary and nonredundant with peroxidase for lignin polymerization during vascular development in Arabidopsis. Plant Cell **25,** 3976-3987.

Zhong, R., and Ye, Z.H. (2014). Complexity of the transcriptional network controlling secondary wall biosynthesis. Plant Sci **229,** 193-207.

Zhong, R., Demura, T., and Ye, Z.H. (2006). SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of Arabidopsis. Plant Cell **18,** 3158-3170.

Zhou, J., Zhong, R., and Ye, Z.H. (2014). Arabidopsis NAC domain proteins, VND1 to VND5, are transcriptional regulators of secondary wall biosynthesis in vessels. PLoS One **9,** e105726.